

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS

RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF

N.V. PHILIPS' GLOEILAMPENFABRIEKEN

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, HOLLAND

A NEW METHOD OF COUNTERACTING NOISE IN SOUND-FILM REPRODUCTION

by W. K. WESTMIJZE.

534.862.4

Reproduction of the sound recorded on sound-film is usually accomplished by means of a narrow beam of light thrown upon the film in a direction perpendicular to that in which the sound track is moving. The fluctuations in the light flux passed through are converted into sound. With this method a noise results which is caused by the fact that part of the light passed through is intercepted by specks of dust, scratches, etc. on the sound track, especially when the film has already been used several times. This article describes a method of counteracting this noise in cases where the sound is recorded as so-called amplitude writing. The beam of light is replaced by a series of equidistant light spots moving with great velocity perpendicular to the sound track. In addition to the theoretical fundamentals of the method, a practical form of application is also discussed.

The ordinary method of reproduction

The reproduction of sound recorded on sound film is usually reproduced in the following manner. A narrow beam of light is thrown on the film perpendicular to its direction of motion. Confining ourselves to the case where the sound is recorded as so-called amplitude-writing, such as for example with the Philips-Miller film¹⁾, the quantity of light passing through the film depends upon the width of the sound track (and of course of the beam). The light passes through to a photocell and is converted into an electric current which may be considered as a direct current upon which an alternating current is superposed. The magnitude of this direct current depends upon the width of the so-called zero track, i.e. the track which is made when no sound vibrations are being recorded. The zero track is unavoidable, since otherwise modulation would be impossible. It is easy to understand that its width must be equal at least to once or twice the maximum modulation amplitude, according as the modulation takes place on one side or on both sides of the track.

The alternating current depends upon the modulation of the track and thus on the sound vibrations recorded, and if the light beam were infinitesimally narrow the trend of this current would be an exact copy of the sound vibrations. Actually the beam

has a finite width Δ , but even so the relation between the sound vibrations recorded and the corresponding vibrations of the light flux can easily be determined. Let us assume that the sound track is modulated by one harmonic vibration. Such a vibration is represented in fig. 1. When

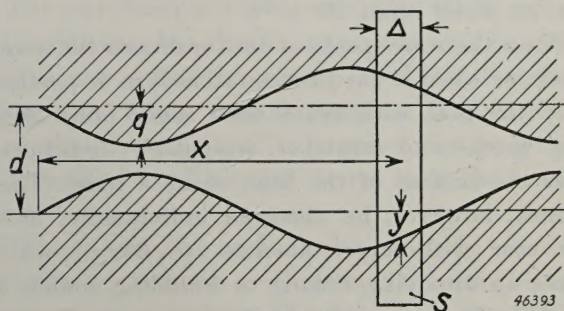


Fig. 1. Diagram of the usual method of scanning. The film with the modulated sound track travels past this beam S. The variations in the light flux passed through are registered by a photocell. In the diagram the track is modulated on both sides by a purely sinusoidal vibration. d width of the unmodulated track, q amplitude of the vibration with which the track is modulated, y depth of modulation at the point with the abscis x , Δ width of the slit.

this vibration corresponds to a tone of ν oscillations per second, and when the velocity at which the film is travelling is v cm/sec, there are ν/v vibrations per cm of film and the vibration can be represented by the equation $y = q \cos 2\pi \nu/v \cdot x$, where y is the depth of modulation and x the length of film passed, measured from an arbitrary zero point,

¹⁾ For a description of the Philips Miller system, see Philips Techn. Rev. 1, 107, 135, 211, 1936.

The amount of light passed through is then proportional to

$$\Delta \cdot d + 2 \int_{x^{-1/2}\Delta}^{x+1/2\Delta} q \cos(2\pi \frac{v}{v} \xi) d\xi = \Delta \cdot d +$$

$$+ \Delta q \cdot \frac{\sin \pi \frac{v}{v} \Delta}{\pi \frac{v}{v} \Delta} \cos 2\pi \frac{v}{v} x.$$

d representing the width of the zero track. From the result it is immediately clear that a DC and an AC component are present, while it is also clear that the amplitude of the AC component is multiplied by a factor which depends upon the frequency v . This factor

$$\frac{\sin \pi \frac{v}{v} \Delta}{\pi \frac{v}{v} \Delta}$$

is equal to unity when $v = 0$, and then decreases. In order that the highest frequencies to be reproduced should not be attenuated by more than about 3.5 db compared with the lowest (such an attenuation is still permissible) it is necessary that

$$\pi \cdot \frac{v_{max}}{v} \cdot \Delta < 1.5.$$

With $v_{max} = 8000$ and $v = 32$ cm/sec this results in $\Delta < 0.002$ cm. The light-beam may therefore not be wider than 20 μ .²⁾

When there are specks of dust or dirt on the sound track or when it has been scratched, as is particularly the case with much used sound films, these tiny specks and scratches, irregularly distributed over the surface of the film, cause a noise. They cannot, however, be observed individually, as is the case with larger particles ($> 80 \mu$), which cause an annoying ticking or bubbling sound. It would mean a considerable improvement in reproduction if this noise could be avoided.

For some time already a system has been in use which diminishes this noise. It is based on the following principle. The noise is most annoying during the soft passages, i.e. when the depth of modulation is slight. In sound recording it is now arranged, by means of suitable connections, that during these passages the zero track becomes narrower, thus reducing the area upon which the troublesome specks or scratches may occur and thereby also the noise. During the louder passages

the zero track again becomes wider, and thus also the noise becomes louder, but this is less troublesome here because for the greater part it is drowned out by the music or speech.

This method, therefore, does not eliminate the noise, but only reduces it during the soft passages.

Principle of high-frequency scanning

We have seen that in the scanning method described above the noise is caused by contaminations on the transparent part of the film between the two edges of the sound track. This phenomenon therefore also occurs when the edges of the track, which actually represent the sound, are ideal. With the method of high-frequency scanning, about to be discussed, only the edges of the track are scanned; the influence of the part between the edges is eliminated and thus also the noise, in so far as it is caused by specks on the transparent part of the film. Of course the noise resulting from imperfections in the edges of the sound track, to which we shall return later, still remains, just as with the method of zero track adaptation discussed above.

With this method of scanning, instead of a narrow slit of light, we have a series of light spots moving at a very high velocity and at regular intervals perpendicularly across the film. Since the sound track is also moving, the light spots actually move in an oblique direction across the film. Here, too, the light passing through falls on a photoelectric cell, which gives a current impulse during the time that the light spot is moving between the edges of the track. The image of this impulse is approximately rectangular. The intensity of the impulse is determined by the intensity of the beam of light employed. The duration of the impulse depends upon the width of the track at the point where the light spot crosses it. Thus in fig. 2 AB in the lower half corresponds to ab in the upper half, the same being true of CD and cd , etc. It is essential to note that the beginning and end points of the blocks are fixed by the edges of the sound track. (For the sake of clearness the obliqueness of the paths of the light spot across the film is exaggerated.) Contaminations on the film are manifested by variations in the beam of light passed through and consequently the image of the current impulses is not actually as shown in fig. 2b, but as in fig. 3; between A and D the current is not constant, variations occurring of an accidental nature. The great advantage achieved lies, however, in the fact that the disturbances are separated from the phenomenon to be reproduced, the former affecting

²⁾ Cf. J. F. Schouten, Synthetic Sound, Philips Techn. Rev. 4, 167, 1939.

the height of the blocks, while the latter only affects the beginning and end points of the blocks.

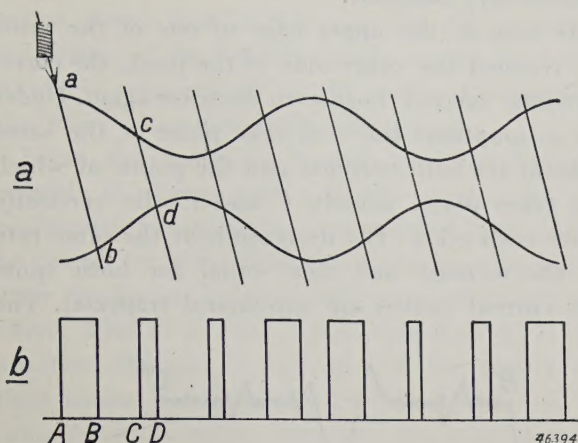


Fig. 2. Diagram of the high-frequency method of scanning. A series of equidistant spots of light travel at a high velocity across the film. Owing to the fact that the film is also traveling at the same time, the light spots describe paths which are oblique with respect to the film and which are given in fig. a. The slope of these paths is very much exaggerated for the sake of clearness. As long as a light spot is inside the edges of the track, a current flows in the photocell. The form of the signal leaving the photocell is shown in fig. b. The block AB corresponds to the path ab, etc.

Therefore the disturbances can easily be eliminated by sending the whole signal through a limiter which only passes signals up to a certain amplitude. In this way the disturbances are, as it were, cut off. For the current variation shown in fig. 3 a limitation to the level EF would be sufficient to bring about this elimination. If the signal is afterwards so amplified that the amplitude is increased in the ratio BA/EA , a signal is obtained which is absolutely identical with what would have been obtained if the sound track had been everywhere uniformly transparent.

We must now consider the question as to how we can derive the original sound frequencies from the block-signal. The frequency spectrum of this

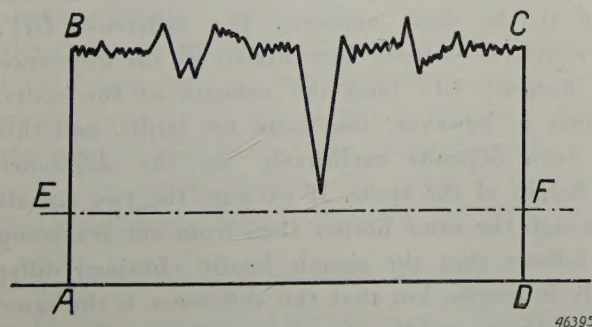


Fig. 3. Image of a current impulse from the photocell. The intensity variations are due to contaminations on the film in the path of the scanning light spot. The essence of the method lies in the fact that the influence of the contaminations can be eliminated by passing the signal through a limiter. Limitation to the level EF would in this case be sufficient.

signal must first be investigated. This involves complicated calculations which will be further dealt with on another occasion. Suffice it here to go into a few qualitative considerations. Let us first examine the unmodulated signal. It consists of congruent blocks having a frequency μ (the scanning frequency). If a Fourier analysis is made of this signal, vibrations with the frequencies μ , 2μ , 3μ , etc. are obtained. If we now modulate the block signal with a frequency ν , secondary frequencies then appear in the spectrum: $\mu \pm \nu$; $\mu \pm 2\nu$; ... $2\mu \pm \nu$; $2\mu \pm 2\nu$; ... $3\mu \pm \nu$; etc. It is, however, quite obvious that also the frequency ν itself will occur. Let us again consider fig. 2. The blocks corresponding to the wide parts of the track are wide and consequently the wide parts give rise to relatively long current impulses with short interruptions. In the case of the narrow parts of the track it is just the reverse. If we now pass this signal through a suitable filter, i.e. a low-pass filter, with limiting frequency coinciding with the highest frequency that has to be passed through, the result is that the signal, roughly speaking, is replaced by a progressive average over a certain time interval approximately of the order of $1/4$ of the time of vibration of the limiting frequency. Thus in each case a number of successive blocks is averaged and the result is a signal which is strong when the blocks are wide and weak when they are narrow, thus an alternating current with a frequency ν corresponding to the frequency of the vibration originally registered.

For reproduction it is essential that the frequency ν should occur but that 2ν , 3ν etc. should be absent. That this is indeed the case is proved by calculation, though it is not easy to imagine. It is obvious, however, that this is of importance, for, as a rule, with ν also 2ν and possibly 3ν etc. lie in the audible range.

We can now also make it clear that the scanning frequency μ must be much greater than the highest frequency ν to be reproduced, because in addition to μ owing to the modulation also the tones $\mu - \nu$, $\mu - 2\nu$, etc. occur. These tones become weaker as we get farther away from the frequency μ .

Calculation shows that the frequency $\mu - 5\nu$ is already 60 db weaker than the frequency ν . The frequency $\mu - 4\nu$ would still be strong enough to be disturbing. If we are to eliminate this by means of a filter, then it must fall outside the audible range, and this means that:

$$\mu - 4\nu_{\max} > \nu_{\max}, \text{ or } \mu > 5\nu_{\max},$$

where $\nu_{\max}$ represents the highest frequency of the audible region which is to be reproduced.

Taking $v_{max} = 8000$ c/sec, it follows that $\mu = 40\,000$ c/sec.

Limitations of the effect of the method

It must be pointed out that not all disturbances can be eliminated in the manner described. Two cases must be examined separately.

In the first place a contamination may be so large and consequently intercept so much light as to cause the photocurrent to fall below the value that passes through the limiter. The result is a "dent" in the corresponding block which again causes a disturbance. This is especially the case when the light is entirely cut off by the contamination, in which event one light spot produces two current impulses (blocks). However, by giving the light spot an oblong shape it is possible to ensure that this case seldom occurs. Already in the beginning of this article it was observed that the width of the light beam in the ordinary method of scanning may not be more than $20\ \mu$, because otherwise the high tones would be weakened. This applies also for the width of the light spots, but not for their height. (By width we mean here the dimension perpendicular to the motion of the light spots and by height the dimension parallel to that motion.) An increase in the height for instance to $100\ \mu$ has by first approximation the same effect on the fluctuations of the transmitted light flux as if the zero track had been taken $100 - 20 = 80\ \mu$ wider and the height left the same. This can easily be explained: Owing to the finite height of the light spots the photocurrent impulses do not have the form of rectangles (apart from the disturbances due to contaminations), but of equilateral trapezia. During the time that the light spot is moving over the edge of the track, the intensity increases from zero to the maximum value and decreases again from the maximum to zero. In fig. 4 two cases are depicted for different heights of the light spots. It is assumed that they begin to pass over the track at the same moment. The photocurrent impulses then begin at the same moment for both, at the point *A*. We further assume, of course, that the two light spots move at the same velocity, so that the light intensity increases in the same way and the trend of the photocurrent will be the same in both cases, for instance along *AB*. A difference occurs only when the lowest light spot is completely over the track, let us say at *B*. From that moment the corresponding photocurrent (except for disturbances) remains constant. For a short while, however, the current corresponding to the highest spot continues to increase at the same rate, until this

spot is also entirely over the track, let us say at *B'*, from which moment the second current, too, is (practically) constant.

As soon as the upper edge of one of the spots has reached the other side of the track, the corresponding current begins to decrease again. Under our assumptions this will take place at the same moment for both currents and the points at which this takes place, namely *C* and *C'*, lie vertically above each other. The decrease is at the same rate as the increase and thus equal for both spots (the current curves are equilateral trapezia). The

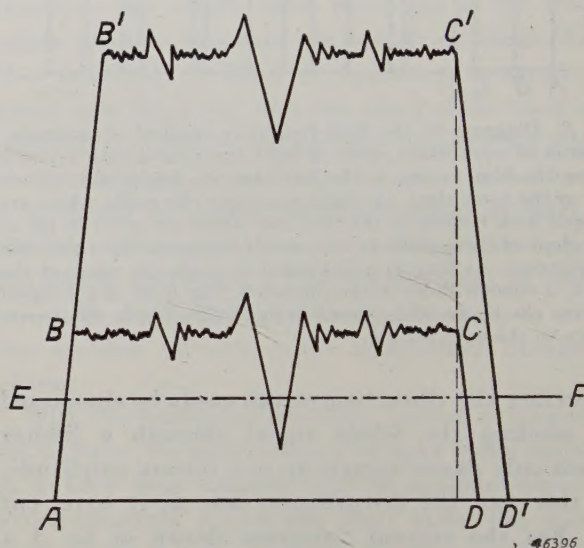


Fig. 4. Influence of the height of the light spot on the form of the photocurrent impulses excited by the light spot. Owing to the fact that the spot has a finite height, some time elapses before the whole spot is over the track. During that time the current increases continually. The impulse *ABCD* is due to a low spot, the impulse *AB'C'D'* to a higher one, the top-side of both spots having reached the edge of the track at the same moment. In the second case the average current of the photosignal is larger. Limitation of the signal to the level *EF* is therefore sufficient to eliminate all disturbances in the second case but not in the first case.

currents thus decrease according to two parallel straight lines, $CD \parallel C'D'$. Therefore they do not end at the same moment. The difference DD' , however, is entirely determined by the difference in intensity CC' (and the velocity of the spots, which is however, the same for both), and this in turn depends exclusively on the difference in height of the spots. If we pass the two signals through the same limiter then from our reasoning it follows that the signals finally obtained differ only in length, but that this difference is the same for all blocks and therefore has no effect on the sound to be ultimately reproduced. It only alters the *DC* component of the photocurrent signal, just as a change in the width of the zero track would do, and this is suppressed by a filter. If the height of

the spot is greater than the width of the track, the situation is somewhat different, but a closer investigation shows that in this case too the length of the blocks of the limited signal, except for a constant, is proportional to the width of the track at the place where the light spot passed.

From the foregoing it will be clear that it is possible to choose such a height of the spot that practically speaking the transmitted light cannot be cut off by contaminations to such a degree that after limitation such disturbances still have any effect. This is in fact demonstrated in fig. 4. The absolute changes in intensity of the transmitted light beams resulting from contaminations are the same for both spots. Therefore the noise assumed to be present in this case is without influence on the limited signal with the higher spot, but with the lower spot it does leave a disturbance in the limited signal.

to vibrate with respect to a diaphragm. In both cases we may consider the vibration as being brought about with a moving light source and a stationary optical system, but also with a stationary light source and a moving optical system. Finally the vibrations may be construed as being brought about by electrical means as well as by mechanical means. We shall here confine ourselves to the description of a method worked out by us in which the scanning is accomplished with a moving light spot obtained from a mechanically moved optical system.

In fig. 5 a diagram is given of the arrangement employed. The light from a linear source is projected by a lens several mm from the edge of a disc which can be rapidly rotated. In this disc radial slits have been sawed beginning at the edge. When the disc is rotating rapidly, therefore, each slit allows a fraction of the light from the image to pass through. The image of the illuminated opening is

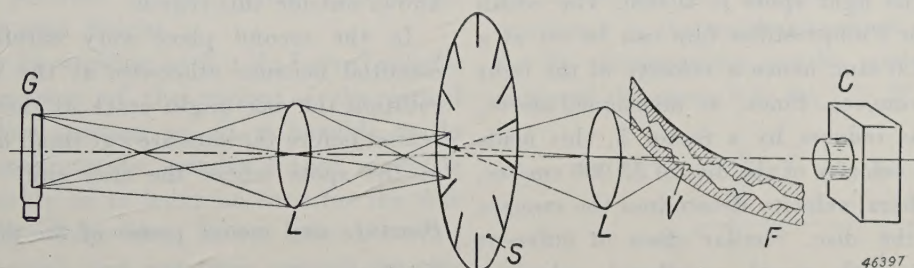


Fig. 5. Diagram of the set-up for high-frequency scanning. *G* source of light with linear filament. *L* lenses. *S* rotating disc with slits. *F* film with modulated sound track. *V* light spot. *C* photocell.

The second possibility of disturbances occurs when a contamination lies exactly tangent to or across the edge of the track. This alters the form of the limitation. The disturbance caused by such an imperfection in the edge of the track is not eliminated by the method discussed here. The chance of such a disturbance occurring, however, is slight compared with that caused by a speck elsewhere on the track. The modulated track is at least 1 mm wide, so that the chance of contaminations, even of the size of 100 μ , coming to lie at the edge is only 20%; most contaminations, however, are much smaller and there is therefore still less chance of their lying at the edge of the track.

One possible construction of the apparatus

The above-described high-frequency scanning can be realized in different ways. In the first place the sound track can be scanned by a moving light spot, as has been assumed in the foregoing. In principle the same results can be attained by projecting the image of the sound track and causing this image

focused on the sound film by means of a second lens. The light passed through the sound track falls on a photocell and gives rise to the photocurrents already mentioned.

The practical realization of such a set-up involves a number of technical difficulties which we shall now discuss.

The choice of light source

We have already remarked that the width of the light spot on the film may not amount to more than 20 μ . Furthermore it must be very sharp (the transition from light to dark must take place within a distance of not more than a few μ) and not only when the projection is along the axis of the system, but also when the image is about 1 mm above or below it. Finally the light must be of sufficient intensity to excite a reasonably amplifiable photocurrent. These conditions make certain demands on the optical system and the source of light.

Linear light sources whose incandescent body is narrower than 80 μ are difficult to produce.

This implies that the optical system must be a reducing one. The same conclusion is reached from the requirement of sharpness of projection. A five-fold reduction suffices for both requirements. This reduction is mainly effected by the second lens. The first lens gives practically an image of 1:1. The requirements for sharpness of the image make it necessary to work with small opening angles.

Finally from the minimum required light intensity of the beam that falls upon the photocell and from the dimensions of the optical system it is to be deduced that the brightness of the light source employed must be at least one thousand candle power per cm^2 . In order to satisfy these requirements a special lamp was constructed.

Construction of the rotating disc

The greatest difficulty lay in the construction of the disc. As already mentioned, the required frequency of the light spots is 40 000. The width of the track for Philips-Miller film can be set at a maximum of 1.6 mm, hence a velocity of the light spots of 6400 cm/sec. Since, as mentioned above, the second lens reduces by a factor 5, this leads to a peripheral velocity of the disc of 32 000 cm/sec. Now the peripheral velocity determines the stresses occurring in the disc. Similar discs of different diameters but with equal peripheral velocities exhibit exactly the same stresses at corresponding points. At a velocity of 32 000 cm/sec. these stresses are enormous and approach the yield point. It is clear that this sets an upper limit for the velocity. In fact if this limit is reached the disc flies to pieces.

Since for different materials under otherwise similar conditions the stresses are proportional to the specific weights, a material had to be found with the most favourable ratio of yield point to specific weight. Moreover, having regard to the motive power for the disc, the material had to be electrically conductive, so that practically only duraluminium and electron could be considered. Furthermore, since the highest stresses occur where the hole is drilled for the spindle, the disc was given a very slightly conical profile.

It can then be calculated that both for duraluminium and for electron the maximum stresses occurring, even at a velocity of 40 000 cm/sec, still remain below half the yield point value. This was in fact confirmed experimentally by investigating at what peripheral velocity a test disc flew to pieces. This was found to be at 60 000 cm/sec (the stresses are proportional to the square of the velocity). From *fig. 6*, which is a photograph of the fragments of

the disc, it is apparent that the break began at the spindle, as was to be expected.

Furthermore, as it was desirable not to make the apparatus too cumbersome, the disc could not be made too large. Its radius was therefore fixed at 5 cm. This means that a speed of rotation of $32\,000/10\pi = 1000$ rev. per sec. is required. Since the slits have to be about $5 \times 1.6 = 8$ mm apart, $10\pi/0.8 = \text{approx. } 40$ slits can be made on such a disc. They are 0.6 mm wide and 3.5 mm long (from this it follows that the length of the light spots on the film is $120\ \mu$). The cutting of the slits requires much care.

In the first place they have to be spaced at exactly equal distances and must be exactly alike, as otherwise the frequency of revolution of the disc appears in the frequency spectrum, and since this lies in the audible region there will be a whistling tone in the sound reproduced. The scanning frequency, which is 40 times as high, lies, as we know, outside this region.

In the second place very careful finishing is essential because otherwise at the high speeds of rotation the disc might crack at the slits. For that reason before the slits are cut small holes are drilled at the spots where the slits end.

Bearings and motive power of the disc.

With the above mentioned very high number of

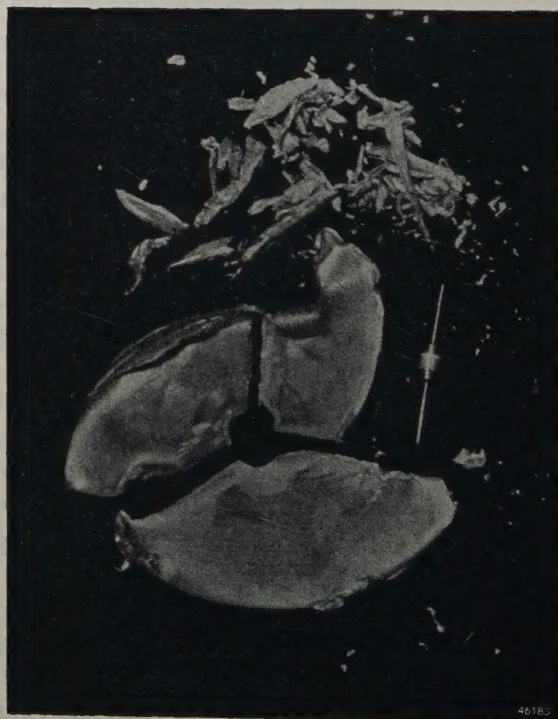


Fig. 6. Fragments of a rotating disc which flew to pieces at a peripheral velocity of about 60 000 cm/sec. The photo shows that the fracture began where the spindle passed through.

revolutions of 1000 per sec special demands are of course made of the bearings. Even a slight eccentricity of the centre of gravity of the disc with respect to the centre of the bearings gives rise to enormous centrifugal forces as the speed increases, resulting in high pressures on the bearings, vibration of the motor, high friction and heavy wear. In order to avoid this the principle of the de Laval shaft was employed, with a thin flexible spindle instead of the usual rigid shaft. Due to the centrifugal force the spindle will sag already at a low number of revolutions, and this sag becomes greater as the speed of rotation increases. When a certain speed is reached, the so-called critical speed, the sag will theoretically even be infinite. Above that speed the sag decreases rapidly and at the limit for infinitely high speed the disc will rotate about its centre of gravity. When this state is reached the sag of the spindle and consequently the pressure on the bearings is very small. The bearing pressure is then mainly determined by the disc's own weight.

A difficulty in working with a de Laval spindle lies in the passing of the region of the critical frequency when starting up. It is possible to do so without breaking the spindle if that region is passed so quickly as to leave no time for the disc to assume large deflections. In our case, however, the driving couple was not large enough for this and we therefore decided to suppress the dangerously large deviations by applying a suitable damping arrangement to the spindle. For that purpose the spindle is passed through eyelets at a short distance from the disc on either side. These eyelets are connected by rods to small pistons moving up and down with a little play in small cylinders containing oil. By this means the lateral movements of the disc are damped, and by choosing suitable dimensions for this device the vibrations in the critical region can be kept sufficiently low. Once the critical region is passed, the disc runs very quietly and speeds of 1000 and 2000 revs/sec are easily attainable.

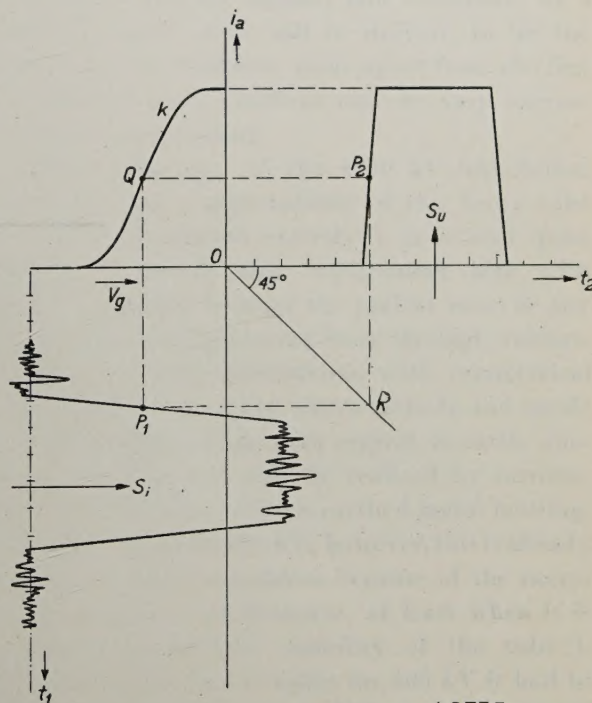
As already remarked in passing, the disc is driven electrically. It is placed in the field of two mutually perpendicular magnetic circuits activated by alternating currents with a frequency of 1500 c/sec and shifted 90° in phase with respect to each other. Each circuit consists of two pole shoes, between which air gaps of about $1\frac{1}{2}$ cm have been cut. The disc is placed in these air gaps. The combination of the two alternating magnetic fields produces a rotating field which turns the disc — made of a conducting material especially for this purpose —

and is able to give it sufficient velocity. In order to minimize friction the disc with the complete driving mechanism is placed in an air-tight housing, so that it can function in a vacuum.

The limitation of the signal and its conversion into sound

The current impulses from the photocell, which are of the order of 10^{-7} A, are first very strongly amplified. For this purpose a wide-band amplifier is used which gives amplification constant within 6 db in a region from 30 to 500 000 c/sec. These voltage impulses are modulated, in the first place by fluctuations resulting from contaminations on the sound film, but in addition a noise connected with the powerful amplification is superposed on the whole signal.

As has already been mentioned in discussing the principle of the method, these disturbances are eliminated by limiting the signal. For this purpose a pentode with high anode resistance is used. As is known, by introducing a sufficiently high resistance in the anode circuit of such a valve the I_a-V_g characteristic can be made to assume the shape of the curve k in fig. 7³⁾. If, then, we apply



46778

Fig. 7. Diagram of the double limitation of the photocurrent signal by means of a pentode. S_i incoming signal showing disturbances caused by contaminations on the film and disturbances due to the powerful amplification. S_u outgoing signal. k I_a-V_g characteristic of the pentode. Starting from an arbitrary point P_1 of S_u the corresponding point P_2 of S_u can be constructed with the aid of the auxiliary points Q and R . Since the time units on the t_1 and t_2 axes are similar, OR cuts the angle between the t axes through the centre.

³⁾ Cf. also Philips techn. Rev. 5, 61, 1940.

to the valve a negative grid voltage so high that $I_a = 0$ even for the most powerful disturbances occurring, in the absence of a signal, and make provision for the signal, on the other hand, to be so powerful as always to generate the maximum anode current, likewise for the most powerful disturbances, then the object has been attained (see fig. 7).

Finally the signal prepared in this manner needs only to be sent through a filter that allows the frequencies of the audible region to pass through and eliminates all the others. It may then be fed to the loudspeaker *via* a power pentode.

Conclusion

By means of the method of counteracting noise described here it is possible to obtain a perceptible improvement in quality of the sound reproduced. At the present stage of development the improvement in the case of new films, which are therefore practically free of contamination, is of no significance. In the case of films which have been used several times, however, the improvement is considerable. The method described thus makes it possible to use a film much longer than was previously possible, with retention of the original quality.

A 400 KILOVOLT INSTALLATION FOR X-RAY THERAPY

by W. HONDIUS BOLDINGH and W. J. OOSTERKAMP.

621.386.1 : 615.849

A description is given of a specially constructed therapy installation for 400 kV, 10 mA, built a few years ago for the Academic Hospital in Groningen. In contrast to the construction usually employed for this voltage, in this case the anode of the X-ray tube is earthed. The focus, which is the source of the X-rays, is situated at the end of a long, earthed metal tube which projects from the high-voltage chamber through a partition into the irradiation chamber. In this way absolute protection against the high voltage is ensured for doctor and patient, while the necessary adjustability of the beam of X-rays is provided for by making the X-ray tube movable and rotatable. In this exceptional arrangement the cathode and the focus of the tube are about 1.80 m apart. The measures necessary for the focussing of the electron beam at such a great distance, as well as several other particulars of the installation, are discussed in detail.

Installations for X-ray depth therapy in hospitals are usually for a voltage of about 200 kV, but higher voltages are now no longer exceptions. In fact Philips have already constructed several installations for 400 kV and higher, such as for example the 1000 kV one installed in the Antoni van Leeuwenhoekhuis in Amsterdam; which was described several years ago in this periodical¹⁾. The medical value of such an installation can by no means be measured simply by the value of the tube voltage applied. The increase in the so-called depth quotient (ratio of the dose at a certain depth under the skin to that on the skin), which was the original reason for raising the voltage, is of course important¹⁾. But, in addition to a large depth quotient, there are also other properties which are of importance for therapeutic uses, such, for example, as the adjustability of the X-ray tube and thus of the beam of X-rays. It is the wish of the doctor to be able to determine quite freely the direction from which the X-rays fall upon the patient. This is especially important in the application of the methods of cross or rotation irradiation, where the X-ray dose necessary for destroying a tumour in the body is divided into several portions, which are applied from different directions in order to spare the healthy tissue and the skin above the tumour as much as possible, thus with the same object as that which holds in the attempt to attain a large depth quotient. The higher the voltage is chosen, the larger the apparatus and the more difficult it is to provide for the adjustability of the X-ray tube. In the case of the 1000 kV installation in Amsterdam, where the object was not only to attain a large depth quotient, but also to investigate possible biological effects of extremely hard X-rays, the possibility of adjustment of the

X-ray tube had to be entirely abandoned and consequently the doctor has at his disposal only a fixed X-ray beam, which emerges from the floor of the irradiation chamber and only the diameter of which can be varied by means of diaphragms. He must place his patient as well as he can with respect to this beam. Cross irradiation is only possible by turning the patient, which meets with objections sometimes of a medical nature (the dropping of internal organs) and sometimes of a practical nature, as it will be difficult to fix the patient in some positions, quite apart from the fact that the necessary positions may be very uncomfortable for the patient.

While in the case of the 1000 kV installation referred to above adjustability of the X-ray tube had to be abandoned entirely²⁾, it is still quite difficult to satisfy this requirement also with 400 kV, especially because the patient must in any case be absolutely protected from the high voltage. In the usual tube construction with symmetrical voltage distribution, thus where cathode and anode are each at half voltage with respect to earth, adequate protection can only be realized by surrounding the entire tube with an earthed metal housing. At a tube voltage of 400 kV, however, this is already found to be quite a problem because of the necessary large insulation distances, at least when it is desired to retain easy mobility of the tube³⁾. Therefore in the case of tubes for 400 kV it had to suffice, in general, to set up light screens of insu-

²⁾ In the meantime new constructions have been developed which ensure a certain degree of mobility and adjustability also with a 1000 kV X-ray tube.

³⁾ In this and other respects the insulated flexible cables connecting the tube with the high voltage generator constitute a particular difficulty. In some constructions the necessity of these cables is avoided by arranging the tube and generator in one assembly unit, but then the heavier weight of such a unit does not allow of ease of mobility.

¹⁾ J. H. van der Tuuk, Philips Techn. Rev. 4, 153, 1939.

lating material to prevent the doctor and the patient from coming dangerously close to the voltage-bearing components.

In the construction of the special 400 kV installation supplied by Philips in 1941 for the Academic Hospital in Groningen, a different method has been followed. The X-ray tube is not symmetrical with respect to voltage distribution, but has an

of the X-rays) can now be adequately protected against the high voltages while the necessary manoeuvrability of the tube is retained. At the anode end the X-ray tube is extended in the form of an earthed metal tube at the end of which is situated the tungsten lozenge that emits the X-rays. The tube is passed through a partition in the wall of the irradiation chamber in such a way

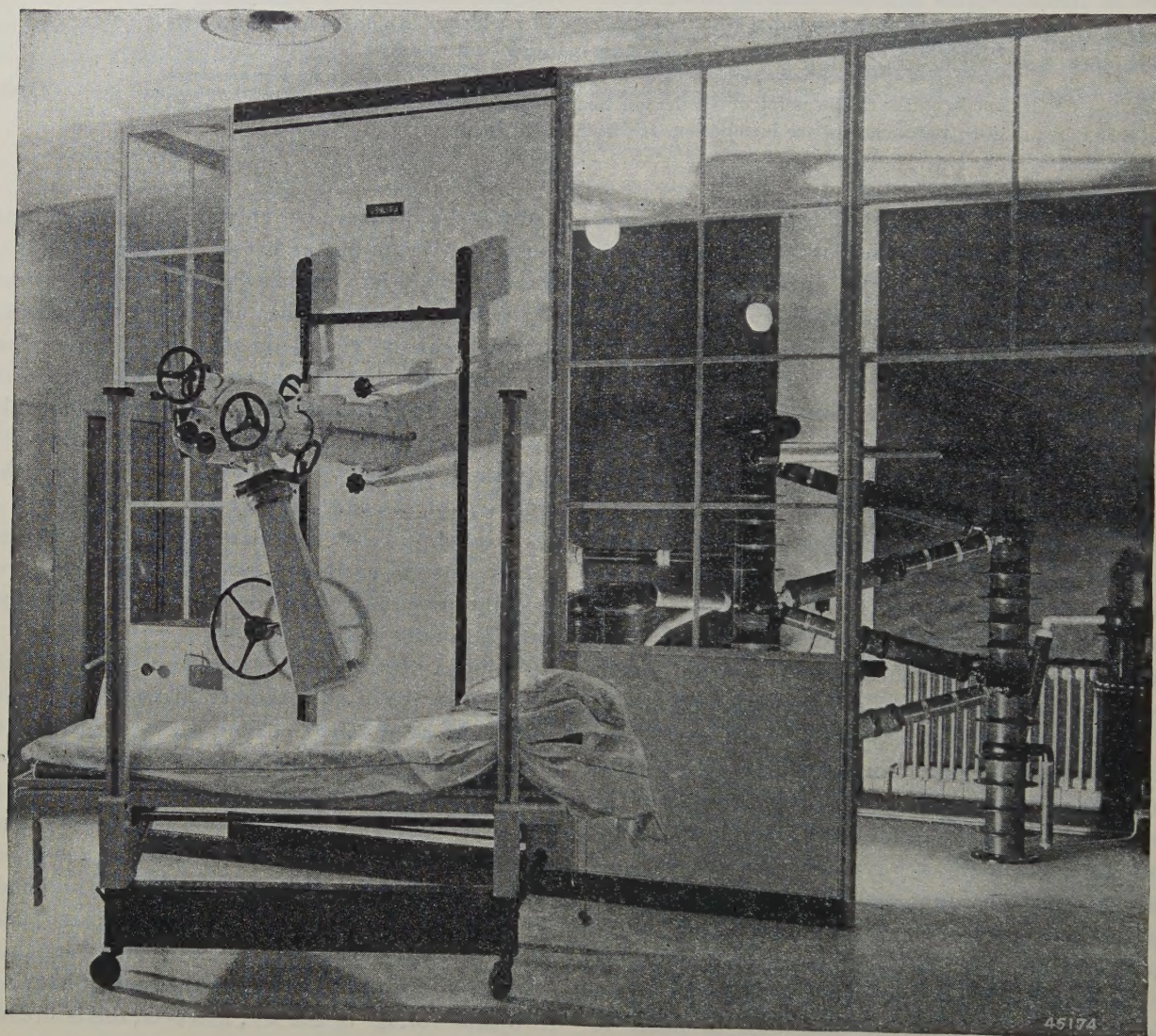


Fig. 1. General view of the therapy installation. On the left may be seen the portion of the X-ray tube projecting into the irradiation chamber, with an irradiation tube affixed to it. In the chamber behind the irradiation chamber is the cascade generator for the excitation of the voltage of 400 kV.

earthed anode. The cathode must then, of course, be brought to a potential of -400 kV with respect to earth, and this, *inter alia*, makes the insulation less simple than when there are two poles at 200 kV with respect to earth, but on the other hand there are the following advantages: 1) the cooling of the anode is very much simplified (this will be discussed later) and 2) the operator, who is only concerned with the anode end of the tube (the actual source

that only the earthed anode tube projects into this chamber, while the cathode end is in the neighbouring room, in which the high-voltage generator is set up. Thus in the irradiation chamber there are no components under high voltage.

This arrangement is shown in *fig. 1*.

By making the metal partition in the wall in which the tube is placed movable vertically and at the same time providing that the tube can be

rotated around its longitudinal axis, sufficient possibilities of adjustment are obtained to be able to approach a patient on a wheeled bed from all directions. It is only necessary that the focus from which the X-rays are emitted should be sufficiently far away from the wall, at least half the length of the bed.

The remarkable consequence of this is that the cathode and the tungsten lozenge, which in an ordinary 400 kV tube are about 8 cm apart, are here about 1.80 m apart. Special means are therefore necessary to focus the electron beam emitted by the cathode on the lozenge. The focussing takes place in three steps, by means of three "lenses", which can be seen in *fig. 2*. An electrostatic lens (the cathode cap) concentrates the electrons on the entrance to the earthed anode tube. A magnetic lens (magnetic coil) at the beginning of the tube concentrates and directs the beam in such a way that on its passage through it does not touch the walls of the tube. A second magnetic coil 30 cm in front of the lozenge finally serves to give the beam the necessary convergence to form a focus of the desired size.

For the first magnetic coil the direction of the

requires some explanation. The size of the focus is determined by two requirements: it may not be so small that at a given power of the tube the specific loading of the tungsten lozenge and consequently its temperature becomes too high; on the other hand the focus should not be made larger than necessary. The edges of the diaphragm through which the cone of X-rays passes out of the tube cast a penumbra which may make a correct dosage at the edges of the field of irradiation difficult, and the width of this penumbra is proportional to the dimensions of the focus. The desired size of focus is thus fixed fairly accurately, and with it the desired diameter of the electron beam on the lozenge. For this diameter the absolute tolerances are even smaller than for the (apparent) width of focus, since the lozenge is at an unusually small angle (30°) to the axis of the electron beam. It is therefore clear that with the first magnetic coil alone, at a distance of 1.5 m from the lozenge, it would be practically impossible to regulate the beam diameter with the required precision.

For depth therapy a very high X-ray intensity is desired, since for the sake of a large depth quotient the tube is generally placed at a fairly great distance

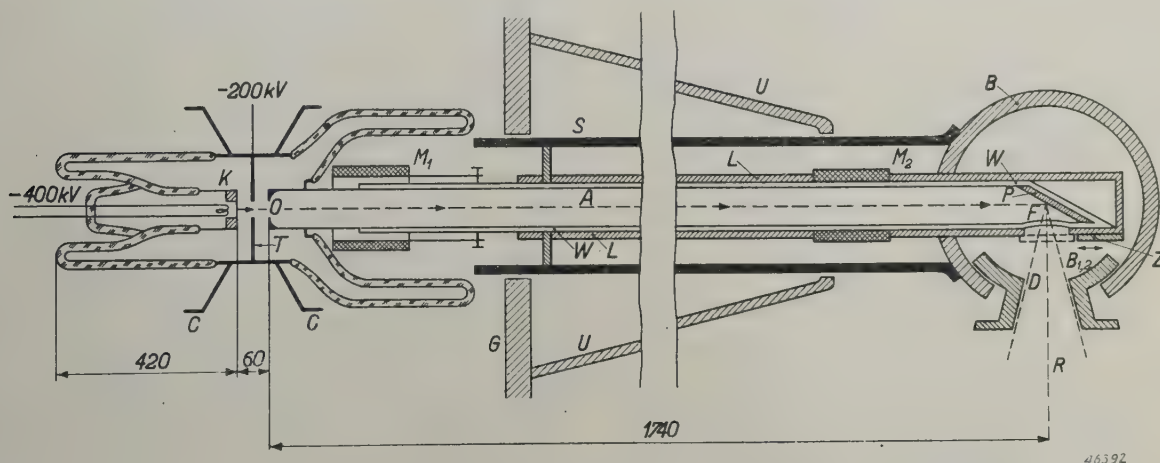


Fig. 2. Simplified diagram of a cross-section of the whole X-ray tube. *K* cathode with focussing cap, *T* partition, *O* opening of the earthed anode tube *A*, *M*₁ first magnetic coil, *M*₂ second magnetic coil. *P* tungsten lozenge on which is the focus *F* which emits the X-rays *R*, *W* cooling jacket with running water from the mains, *L* lead jacket around the anode tube, *B* lead sphere, *B*_{1,2} lead spherical caps with diaphragm openings *D*, *Z* lead slide by means of which the beam can be cut off, *S* steel tube supporting the anode tube and movable in the wall of the room *C*₁ and the collar *U*; *C* corona gaps. For the sake of clearness the case of one irradiation beam only is drawn. Actually there are two exits for the rays in the two spherical caps, one of which is directed forwards and the other backwards with respect to the plane of the drawing. Cf. the photograph.

magnetic field is very critical. This coil is therefore mounted on the anode tube by means of a ring and four supporting points adjustable by means of screws, so that the position of the coil may be corrected during the assembly of the X-ray tube.

The significance of the second magnetic coil

from the patient. Now in order to use a high power in the tube without it being necessary to enlarge the surface of the focus too much, it is important that the specific loading capacity of the lozenge should be high. Owing to the fact that the anode is earthed in the installation in question it was

relatively easy to attain this high loading capacity, because the anode could be provided with water cooling connected to the water mains direct. With a focus of 12×6 mm it was thus possible to increase the power converted into heat in the anode to 4 kW (10 mA at 400 kV DC voltage ⁴).

The cooling water is conducted not only along the tungsten lozenge, but also along the whole anode tube. In spite of the focussing, 10-20 percent of the primary electrons, as well as the secondary electrons from the lozenge strike the walls of the tube, so that a power of about 1 kW converted into heat has to be dissipated from the anode tube. At the same time the cooling water must keep the magnetic coils cool, since heating of these coils may change the coil resistance and with it the current through the coil and thus the intensity of the magnetic field.

The fact, noted above, that the lozenge is placed at a very small angle to the electron beam is due to the desire to be able to treat two patients simultaneously. In order to be able to irradiate each of the two patients from any direction in spite of the mutual hindrance of the two beds, the axes of the two X-ray cones to be used must be far enough apart and also capable of being directed otherwise than perpendicular to the axis of the tube, while at the same time the X-rays may not be emitted at too small an angle to the surface of the lozenge. The sketches in *fig. 3* show that it is favourable in this respect to choose the angle of the lozenge to the beam as small as possible. The focus

on the lozenge emits X-rays in all directions. The two narrow cones of X-rays (maximum angle of divergence 30°) are obtained by providing the end of the anode tube with two spherical lead caps each with a diaphragm opening. In order to direct the two cones of rays, each of the spherical caps can be turned through a certain angle about two mutually perpendicular axes by means of two hand-wheels. A lead sphere fixed rigidly to the tube and enveloping the spherical caps is provided with two openings sufficiently large to permit the passage of the effective cones of rays in all desired directions and prevents outward radiation in undesired directions at any position of the spherical caps. The total thickness of lead in all directions is at least 13 mm, which is sufficient to ensure that everywhere outside the irradiation chamber the so-called tolerance intensity (10^{-5} röntgen per second) is not exceeded. For the adjustment of the X-ray beams with respect to the patients and for the accurate limitation of the field irradiated, an irradiation tube of the desired length, for example 1 m, can be fastened over each of the diaphragm apertures. This can be seen in the photograph (*fig. 1*).

In spite of the adjustability of the two cones of X-rays, the degrees of freedom of the tube mentioned in the beginning of this article (rotation and vertical displacement) are not yet sufficient for the simultaneous treatment of two patients; it must at least be possible to place the two patients at different levels relative to the tube, thus the beds must be adjustable in height. In order to facilitate the placing of the two patients in the desired positions, and particularly to avoid having to move the beds about too much several additional, not strictly necessary degrees of freedom are given to the tube: it can be moved horizontally another 20-30 cm not only perpendicular to but also in the direction of its axis. Vertically the tube can be moved 1.70 m, and the possible angle of rotation about its axis is 330° , so that, for example, patients can be treated with ray beams directed upwards from underneath the bed. All these movements, as well as the adjustments of the diaphragms, are effected by turning the various hand-wheels seen in *fig. 1*.

Besides the end of the anode where the tungsten lozenge is situated the anode tube itself also had to be enveloped in lead along most of its length, for the 15 percent (approx.) of the primary electrons reaching the wall of this tube also excite an intense X-radiation. Here, however, the emitting material is not tungsten, as is that of the lozenge,

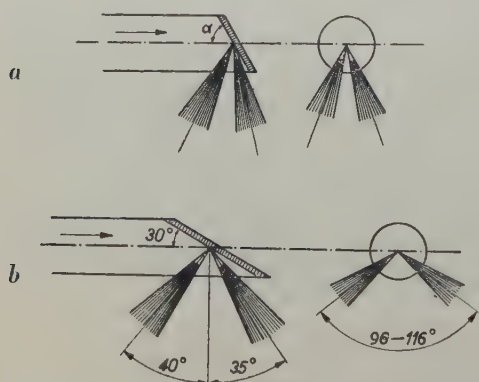


Fig. 3. If the lozenge is at a large angle α to the horizontal electron beam (*a*), the axis of the effective cone of X-rays has only little freedom of motion perpendicular to the beam (right) and still less in forward directions from the perpendicular position (left). When the lozenge is placed at a small angle to the electron beam (*b*) there is much more freedom of movement and two X-ray cones with their axes approximately perpendicular to each other can easily be used at the same time.

⁴) In the construction hitherto customary the maximum load at 400 kV and with the same focus dimensions amounted to only 4 mA. At 400 kV and 10 mA tube current the X-ray intensity with the self-filter of 0.7 mm Cu amounts to 33 r/min at a distance of 1 m from the focus.

but chromium, since the anode tube is made of internally chromium-plated copper. Since the X-ray-emitting property of this metal is much less than that of tungsten (about 30 percent of the emission of tungsten) the lead jacket around the anode tube need only be 6 mm thick. The anode tube with its cooling jacket and lead jacket is supported by a wide steel tube, which also supports the lead sphere and spherical caps on the end of the tube and lends the necessary rigidity to the whole. This steel tube is mounted in the wall partition in such a way as to be rotatable and movable axially.

For the rest the construction of the X-ray tube does not differ very much from what Philips have already supplied for a voltage of 400 kV. Only a few details remain to be mentioned.

In order to obtain the necessary protection against voltage the principle of subdivision of the voltage has been applied ⁵): between the cathode and the beginning of the anode tube a metal partition is introduced which has an aperture for the passage of the primary electron beam and is brought to a potential halfway between that of the cathode and the anode. The high voltage for the tube is furnished by a cascade generator with two stages, which can be seen in fig. 1 in the room at the rear. The necessary voltage of 200 kV for the intermediate electrode of the tube can be drawn off directly between the two stages. The cathode, which is at a potential of 400 kV with respect to earth, is heated by two transformers in series, each with insulation for 200 kV between the primary and secondary windings.

In X-ray tubes for such voltages the various high voltage lead-in points are generally enveloped in large, well rounded metal caps, in order to keep the electric field strength in the space around the tube sufficiently low: the field is "homogenized", *i.e.* an attempt is made to approach as nearly as possible the state of a homogeneous field between two plane parallel electrodes, where the field is smaller than in any other configuration with the same distances. In our case, however, the introduction of large metal caps was undesirable because it would have increased the weight of the already very heavy tube. Therefore the homogenization was effected with the help of a corona gap, *i.e.* toothed metal discs. The high electrical field strength at the teeth of such a disc causes a corona discharge to take place there, which is accompanied by ionization of the air. As a result the air in the vicinity

of the disc edge becomes to a certain extent conductive, and this conducting layer of air has the same action as if a metal conductor were provided there.

In fig. 4 the corona gap on the cathode supply connection of the tube can clearly be seen. The connection of the cathode with the 400 kV terminal of the high-voltage generator must allow for the necessary freedom of motion of the tube in the vertical and in the two horizontal directions, as well as for rotation. A flexible metal tube is used for this purpose. A cap in the form of a long upright open cylinder, which is visible in the centre of fig. 4, provides that the field at the 400 kV terminal remains sufficiently homogenized at all positions of the tube.

The X-ray tube can, if desired, also be used at a lower voltage than 400 kV. For this purpose the initial voltage of the cascade generator has been made adjustable in a large number of steps with the help of an auto-transformer. When the tube voltage is varied, however, the currents of the two magnetic coils serving for the focussing of the electron beam in the anode tube must also be varied. The diffractive action of the magnetic field on the electrons is inversely proportional to their velocity and thus inversely proportional to the square root of the tube voltage (when relativistic corrections are disregarded). The corresponding regulation of the coil currents takes place automatically, since the supply rectifier is coupled with the voltage regulator. The voltage regulator is placed on the operation desk of the installation, where the tube voltage, measured by the current in a resistance in parallel with the tube, can be read off directly on a meter. At the lower voltages the current through the tube can be increased to a maximum of 20 mA, so that the tube can also be fully loaded at a voltage of 200 kV.

On the operation desk there are various auxiliary instruments, such as the indicator of the dosimeter, with which every irradiation is checked. The irradiation of each of the two patients can be begun and ended independently by opening or closing the corresponding "door" of the lead spheres around the lozenge. For this purpose both doors are fitted with two 13 mm thick lead slides which can be operated from the operation desk by two small servo motors mounted at the end of the tube. The position of these slides, open or closed, is indicated by signal lamps. Further there are two series of six signal lamps corresponding to six different ray filters which can be slid in front of each of the two doors. In this way the doctor or

⁵) For a detailed discussion of this now generally accepted principle see the article referred to in footnote 1).

his assistant who is in charge of the irradiation treatment, and who may under no circumstances enter the irradiation chamber during a treatment because of the danger from rays scattered by the

patient, can nevertheless always see whether the correct filter has been inserted and whether the slides have been opened or closed, as the case may be.

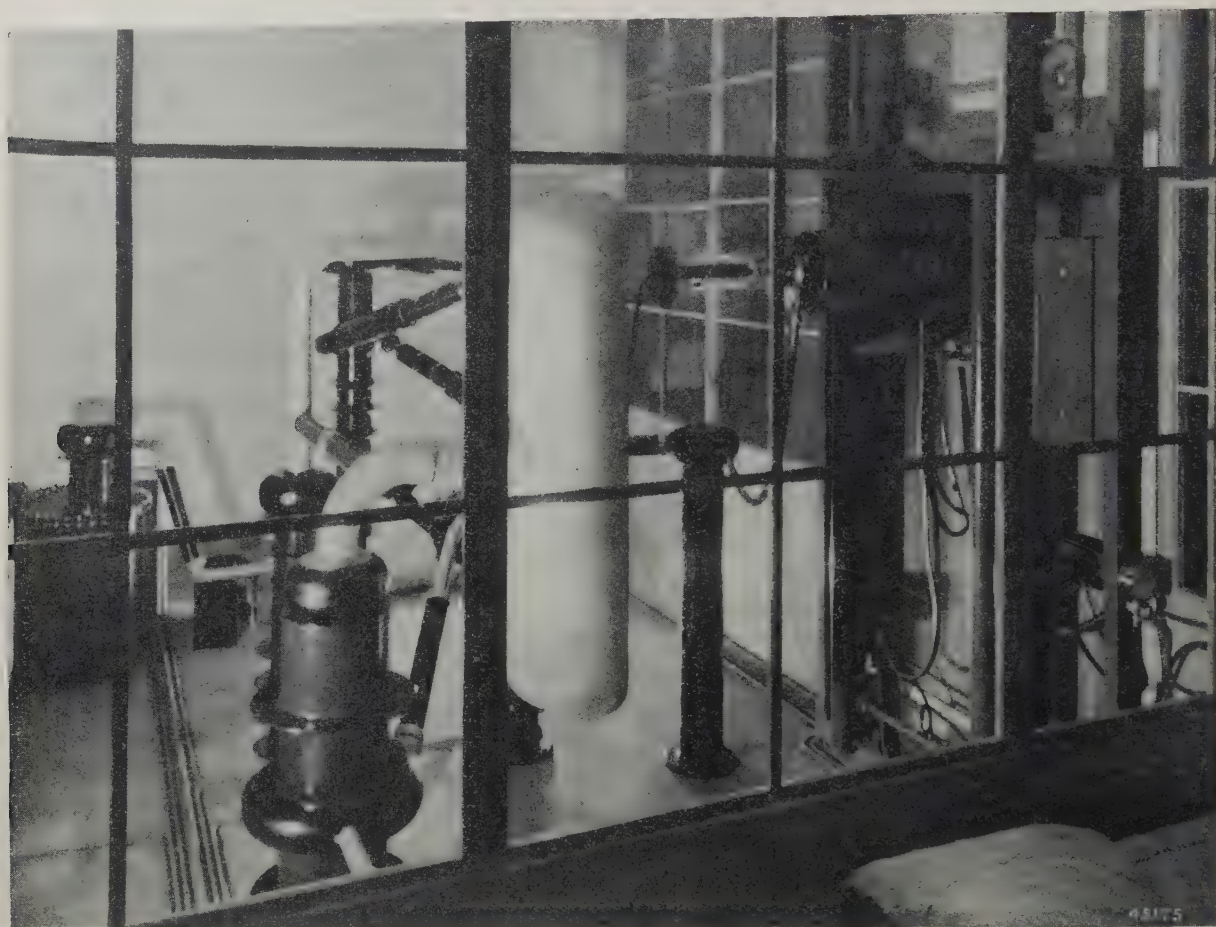


Fig. 4. View of the high-voltage chamber. On the right the end of the X-ray tube projecting through the wall. The voltage supply lead for the cathode can be seen with a corona gap, and in the middle the large hollow electrode cylinder cap for keeping the field homogeneous when the tube is moved horizontally or vertically. To the right of this may also be seen the voltage lead for the intermediate electrode of the tube. To the left in the foreground the transformers for heating the cathode. In the background the cascade generator.

RADIO INVESTIGATION OF THE IONOSPHERE

by C. J. BAKKER.

621.391.11 : 551.510.535

This article gives a survey of investigations made concerning the layers of ionized air (ionosphere) present in the atmosphere high above the earth which are able to reflect radio waves back to the earth, thus making radio reception possible at a great distance beyond the horizon of the transmitter. Various phenomena in shortwave reception, such as fading and the skip distance, are due to the presence of the ionosphere. The density of ions and the so-called "effective" altitude of the ionosphere can be derived from measurements that have been taken. It has been found that particularly the density of ions, and also the reception of short waves, depends to a large extent upon the sun, namely upon its height above the horizon and its eruptive activity. The main, but not the only cause of the ionization is in fact the irradiation of the atmosphere by sunlight, in particular by the part of the spectrum in the far ultra-violet. In the process of ionization this part of the sun's spectrum is absorbed and cannot, therefore, be observed on the earth. Interesting conclusions as to the intensity of the sun's radiation in the far ultra-violet can be drawn from data concerning the ionosphere obtained experimentally by means of reflected radio waves.

After the achievement of the first transatlantic wireless connection by Marconi in 1901, a discussion quickly arose as to how it was possible that the electromagnetic waves sent out could reach the receiving station around the curved surface of the earth. If the propagation of the waves were rectilinear, signals from a transmitter situated beyond the horizon could never be received, so that it had to be assumed that the path of the waves was not a straight line but a curved or a broken line; both hypotheses found supporters.

One year after Marconi's experiments Kennelly and Heaviside, independently of each other, proposed the hypothesis that the high, very rarefied layers of air behave as a mirror for electrical waves and reflect the radio signals back to the earth. The propagation of the waves could then take place along broken lines. The reflective power of these higher layers of air might be caused by an ionization of the air, *i.e.* a splitting of the molecules into electrons and positive ions. Kennelly estimated the lower boundary of the ionization region, the so-called ionosphere, to be at an altitude of 80 km, which, as we now know, is fairly correct, at least as far as the order of magnitude is concerned.

In the meantime attempts had also been made to explain the possibility of reception below the horizon without the hypothesis of a reflecting layer. Various theorists (among whom Rayleigh, Poincaré, later Watson, Sommerfeld, Bremmer and van der Pol¹⁾) showed that the electromagnetic waves are bent along the conducting earth's surface, so that, especially in the case of

long waves, reception far below the horizon is possible.

In the course of further investigations it has been found that this theory of refraction must not be considered as an argument for making the hypothesis of the ionosphere reflections superfluous, but that the two theories supplement each other excellently. In radio communications with waves shorter than 10 m the ionosphere plays no rôle. In the very important wave-length region from 10 to 50 m, on the other hand, the bridging of long distances with relatively slight energy would be impossible without the reflection of the waves against the ionosphere.

Until about 1920 only wave lengths above 200 m were used. After the possibilities of the wave-length region below 200 m had been proved by the bridging of the Atlantic with relatively low power transmitters at about that time, shorter and shorter waves began to be used. Several new phenomena were then encountered which did not occur on long waves or only sporadically, and the cause of these was thought to lie in the presence of the ionosphere. It was found, for example, that reception was generally stronger at night than in the daytime. Another phenomenon was that radio *H*-location on short waves is subject to larger errors at night than in the daytime. To these phenomena could be added that of fading and the fact that with increasing distance reception from an ultra-short wave transmitter does not decrease in intensity continuously, but after having passed a "dead zone" (skip distance) may abruptly attain a greater intensity again. Another interesting phenomenon, and one which is of technical importance, is that the northern lights, the magnetic storms

¹⁾ See: Philips techn. Rev. 4, 245, 1939.

(i.e. great changes in the magnitude and direction of the magnetic field of the earth) and certain kinds of fading are found to be intimately related to each other. Thus, for example, in 1928 it was at first impossible to make wireless contact with the Nobile expedition which had come to grief near Spitsbergen, since the attempts to do so happened to coincide with magnetic storms in the northern lights zone.

The phenomena mentioned have been the subject of profound investigations carried out in many countries, and these have now reached a stage where not only has a fairly complete technical insight been obtained into the causes and the mechanism of many of the phenomena referred to, but also various new fields of geophysical, meteorological and astrophysical investigation have been opened.

Refraction and reflection of radio waves by the ionosphere

A radio wave is an electromagnetic wave which differs from a light wave only in its wave length. It may therefore be expected that the refraction and reflection of radio waves by the ionosphere will in many respects follow the known laws of optics. We shall therefore try to characterize the properties of the ionosphere in the manner customary in optics, by assigning to every point in space a definite index of refraction. What do we mean by this index of refraction? For the case when dissipation can be disregarded it can be derived in a simple way.

According to a familiar relation of Maxwell the index of refraction of a substance n is related to the dielectric constant E by the equation:

$$n^2 = \epsilon \quad (1)$$

We can therefore determine the index of refraction by studying how a gas of the nature of the ionosphere behaves as a dielectric. In doing so we shall consider the ionosphere for the present as a homogeneous gas which in addition to the molecules contains N electrons and N positive ions per cm^3 .

Suppose that in such a gas electrical vibrations are excited with an angular frequency ω and an amplitude E_0 . This electrical alternating field in a dielectric gives rise to electric currents whose density we shall call j . The dielectric constant is then defined by

$$\epsilon = j/j_1,$$

where j_1 represents the displacement current for a vacuum.

The electric current j consists in general of two components, the first of which is the above-men-

tioned displacement current for a vacuum. This is given by:

$$j_1 = \frac{1}{4\pi} \cdot \frac{dE}{dt} \quad (2)$$

The second component j_2 is a convection current, to be ascribed to an actual displacement of charges. If N is the number of charged particles per cm^3 , e the charge and v the velocity, then

$$j_2 = N e v$$

and the dielectric constant is thus:

$$\epsilon = (j_1 + j_2)/j_1 = 1 + \frac{4\pi N e v}{dE/dt} \quad (3)$$

For free electrons such as are present in the ionosphere the velocity v can easily be calculated²⁾. The acceleration of particles with charge e and mass m is determined by the equation

$$m \frac{dv}{dt} = eE. \quad (4)$$

When E changes sinusoidally with an angular frequency ω , then the same is true for v , i.e.

$$\frac{d^2v}{dt^2} = -\omega^2 v.$$

From this and equation (4), after differentiating t , it then follows that:

$$v = -\frac{e dE/dt}{m\omega^2};$$

and by substituting this in (3) one obtains:

$$n^2 = \epsilon = 1 - \frac{4\pi N e^2}{m\omega^2} \quad (5)$$

The apparent dielectric constant is thus less than unity, because the convection current is in opposite phase to the displacement current, and the same is true of the index of refraction. Thus, speaking in optical language, the ionosphere is a less dense medium than empty space.

Now it is known that a wave which passes from an optically dense medium into an optically less dense medium is not only refracted but may also be totally reflected. If n is the index of refraction of the less dense medium, n' that of the denser medium, the condition for total reflection is $n \leq n' \sin a$, where a is the angle of incidence on the

²⁾ In addition to free electrons the ionosphere contains an equally large number of positive ions. Since, however, these possess a mass about 50 000 times as large, they are accelerated to a much smaller extent by electric fields, so that in practice we need only consider as mobile particles the free electrons.

surface of reflection. In our case $n' = 1$. If we now first consider the case where $a = 0$ (signal directed vertically upward), the condition is then $n < 0$, or, according to (5),

$$\frac{4\pi Ne^2}{m\omega^2} \geq 1. \quad (6)$$

For sufficiently low values of ω this condition is always satisfied; low-frequency signals are thus totally reflected by the ionosphere. Above a certain frequency, however, this is no longer the case; for the critical frequency, according to equation (6), the following holds:

$$\frac{4\pi Ne^2}{m\omega_{\text{crit}}^2} = 1, \text{ or } f_{\text{crit}}^2 = \frac{Ne^2}{\pi m}. \quad (7)$$

The critical frequency f_{crit} is thus a measure of the electron density N in the ionosphere.

The signals directed vertically upwards for which we have calculated the critical frequency are of great importance in the study of the ionosphere, as will be discussed below. In practical radio transmission, however, one is usually concerned with waves directed obliquely upwards. The chance of total reflection is thereby increased. If a is the angle of the wave to the vertical then $n < 0 \sin a$ holds for total reflection, or according to equation (5)

$$1 - \frac{4\pi Ne^2}{m\omega^2} \leq \sin^2 a < 1 - \cos^2 a. \quad (8)$$

If we set $a = 0$ we again obtain equation (6) for the critical frequency. If, conversely, we choose any arbitrary frequency f greater than f_{crit} , equation (8) then gives us a critical angle of reflection:

$$\cos^2 a_{\text{crit}} = \frac{4\pi Ne^2}{m\omega^2} = \frac{Ne^2}{\pi m f^2} = \frac{f_{\text{crit}}^2}{f^2}. \quad (9)$$

Reflection only occurs for $a > a_{\text{crit}}$, while waves with an angle of incidence $a < a_{\text{crit}}$ pass through the ionosphere. Practically this amounts to the fact that only at a certain minimum distance from the transmitter can the reflected signals be observed.

A simplified representation of the situation is illustrated by *fig. 1a*, where it is assumed that the altitude of the ionosphere is so small in relation to the radius of the earth that the earth may be considered flat, while, moreover, the reflection from the ionosphere takes place in the form of a sharp angle.

The fact that the reflected radiation only becomes observable at a definite distance r_1 is familiar in practice in the existence of a "dead zone"³⁾ (skip

distance). If one is concerned with case *a* in *fig. 1* the radius of the skip distance can easily be calculated, and one finds:

$$r_1^2 = 4h^2 \left| \frac{\pi f^2 m}{Ne^2} - 1 \right|,$$

where h is the altitude of that layer of the ionosphere from which the radiation is reflected. Since this height may amount to several hundred kilometers, however, in many cases the assumption of a flat plane as the earth's surface is not sufficiently accurate, and consequently the curvature of the earth has to be taken into account. Furthermore, the angle of reflection is not so sharp, since the boundaries of the ionosphere are more or less vague and the deflection is therefore more gradual. Taking both these factors into account, the picture represented in *fig. 1b* is obtained. The region commanded by the radiation reflected at the ionosphere then has not only an inner boundary r_1 (skip distance) but also an outer boundary r_2 , namely where the reflected ray still just touches the earth. Under favourable conditions this distance amounts to about 1700 km for signals reflected by a layer at an altitude of 100 km and about 3600 km for a layer at 300 km. Transmission of short-wave signals over greater distances by single reflection at the ionosphere is therefore impossible. With multiple re-

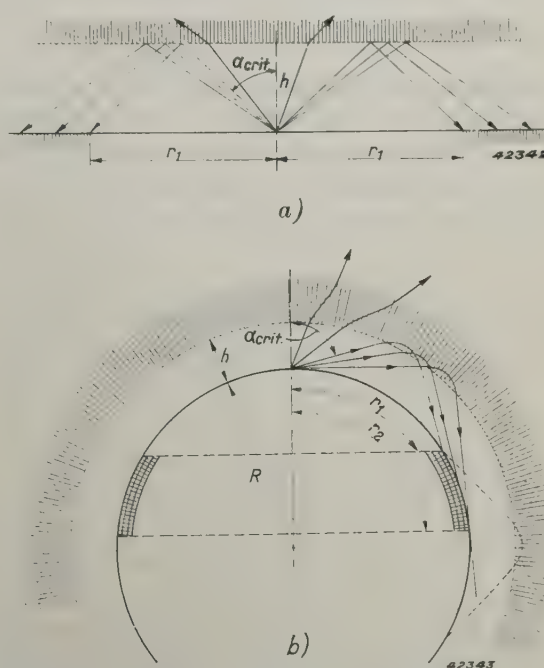


Fig. 1. Reflection of waves by the ionosphere when the earth is considered flat and reflection sharp (a) and when the earth is curved and reflection more gradual (b). Rays with an angle smaller than a_{crit} are not reflected but pass through the ionosphere. The region commanded by the once-reflected waves is limited by a minimum radius r_1 (skip distance) and in the case of the curved surface of the earth also by a maximum radius r_2 .

³⁾ Within the dead zone signals can sometimes be received at irregular times; these are, however, unsuitable for regular communication. The reflection in these cases probably takes place from irregularly occurring electron clouds in or between the normal layers of the ionosphere.

flection at the earth and the ionosphere greater distances can of course be bridged.

The higher the frequency is chosen, the narrower the annular region commanded by the reflected waves, since the radius r_1 of the skip distance zone becomes steadily larger. For the limiting frequency at which it shrinks to nothing one finds about 30 megacycles/sec ($\lambda = 10$ m), depending on the activity of the sun, while f_{crit} lies at about 3 megacycles/sec ($\lambda = 100$ m). In the following we shall go more deeply into the numerical data and also discuss what these empirical values are able to show us about the altitude and condition of the ionosphere.

Altitude and electron density of the ionosphere

Experimental investigation of the ionosphere is being carried out in very many countries, so that in spite of the fact that it has been in progress only some 20 or 30 years there is already considerable empirical material available.

The data of most interest are those concerning the altitude of the ionosphere and the degree of ionization.

Several methods of determining the altitude are in use. A method originally employed by Appleton makes use of interferences between the so-called carrier wave and the radiation reflected by the ionosphere (see fig. 2). The distance between transmitter and receiver chosen is such that the intensity of the carrier wave received is about equal to that of the reflected wave. The two waves will amplify each other when $2(a-a') = m\lambda$ and attenuate each other for $2(a-a') = (m + \frac{1}{2})\lambda$ (m being a whole number and λ the wave length).

If the wave length is decreased continuously

from λ to λ' , amplified and attenuated signals are received alternately. The number of times that an amplified signal is received is:

$$\mu = 2(a-a')/\lambda' - 2(a-a')/\lambda.$$

By determining μ , a , λ and λ' , a' and thus also h' can be calculated. This "effective altitude" h' is greater than the actual altitude of the point where the reflection takes place, since the velocity with which the signal is propagated in the ionosphere is less than the velocity c of light in vacuum.



Fig. 3. The transmitted signal A and the reflected signals B in the determination of the effective altitude of the ionosphere according to the method of Breit and Tuve.

The phenomenon underlying Appleton's method, viz. interference of the waves that have travelled different paths from the transmitter to the receiver, is manifested in normal radio reception as the familiar fading.

Interpretation of the results of the measurements by the Appleton method is rather difficult. Therefore a different method is now usually employed, the "echo method" developed by Breit and Tuve, upon which, as a matter of fact, the now universally known "radar" is based.

By means of a transmitter short wave trains are sent out directly upwards at intervals of, for instance, 1/50 sec (see fig. 3). A receiver not far from the transmitter records the carrier wave and immediately afterwards the echoes due to reflection from the ionosphere. From the time t_e elapsing between the reception of the carrier wave and that of the reflected wave, (echo time), the effective altitude h' of the ionosphere can be calculated by the relation $h' = \frac{1}{2} ct_e$.

Some results obtained by the Breit and Tuve echo method are reproduced in fig. 4. The registogram shows the echo time of the reflected signal as a function of the frequency. It is found that up to about 4 megacycles/sec ($\lambda = 75$ m) an echo time of 60 microseconds occurs, corresponding to an apparent altitude of 100 km. At frequencies above 3.5 megacycles/sec the echo time begins to increase gradually; apparently waves of these frequencies penetrate deeper into the ionosphere. The critical frequency is passed at 4 megacycles/sec. The signal which penetrates through the lowest layer of the ionosphere above this limit does not, however, disappear into space but is reflected from a higher

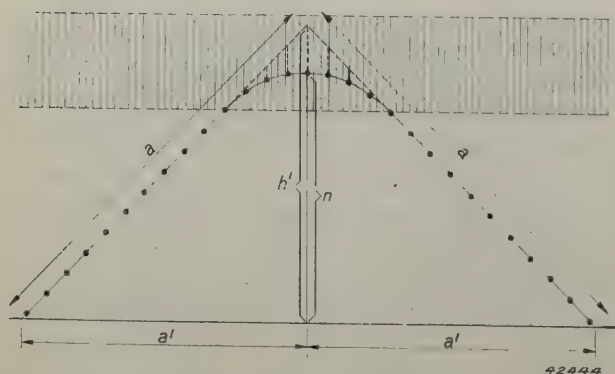


Fig. 2. Determination of the effective altitude h' of the ionosphere from interferences resulting from the difference in path $a-a'$. It may be seen that the actual altitude is smaller since the waves do not experience a sharp reflection but are gradually bent. The dots on the line indicating the path of the electric wave represent time intervals and show that the rounding-off of the angle results in no decrease in the transit time, since in the ionosphere the signal is propagated more slowly.

layer ($h \approx 200$ km). This so-called F_1 layer is much less sharply bounded than the underlying layer, which is indicated as the E layer. This is evident in the diagram from the fact that the altitude at which reflection occurs increases sharply

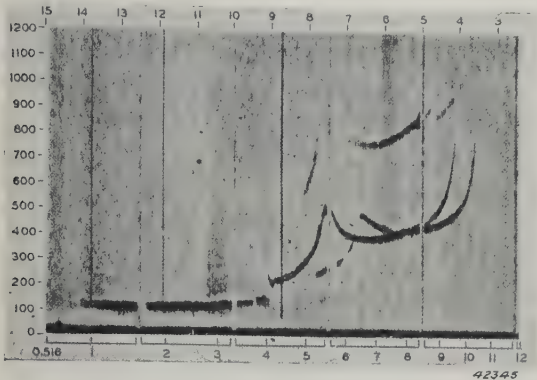


Fig. 4. Recording of the effective altitude of the ionosphere in km as a function of the frequency in megacycles/sec. when the sun is high. Registrogram of the Carnegie Insitution Washington; taken from Darrow, Bell Syst. techn. J. 19, 455, 1940.

with the frequency. At a frequency of 5.5 megacycles/sec the signal penetrates also through the F_1 layer, reflection then occurring at the so-called F_2 layer, which has an altitude of 300 km. Not until a frequency of 10 megacycles/sec is reached are the waves able to penetrate through this layer, when the signal disappears into space.

In the frequency region in which the F_2 layer is effective the diagram shows a second echo, which can be ascribed to waves which have covered the distance between earth and the ionosphere and back twice. Furthermore a remarkable doubling of the diagram lines is visible. This latter effect is connected with the terrestrial magnetic field and will be discussed briefly farther on in this article.

The critical frequencies found for the reflections at the various layers make it possible to calculate the electron density N at the point where the waves are reflected. According to equation (7) this quantity is:

$$N = \frac{\pi m f_{crit}^2}{e^2} = 1.24 \cdot 10^{-8} f_{crit}^2 \quad (10)$$

By the combination of determinations of N with measurements of h' a fairly complete picture can be obtained of the distribution of the electrons over the different layers of the ionosphere.

Results of the investigation

Apart from systematic variations with the season of the year and the position of the sun, the state of the ionosphere is found to exhibit rather great fluctuations. Electron densities occur which are 10

times as great as the average. In general the following conditions can be determined.

At an altitude of about 50 km in the daytime there is a slightly ionized layer ("ozone" or D layer). At night the electron density in the D layer is very slight. The second layer (E layer) is permanent and lies at about 120 km altitude. Above that at 200 and 300 km altitude lie at least two other layers (F_1 and F_2) more or less distinct in the daytime but immediately after sunset merged into one, the F layer; sometimes in the daytime in winter too the separation between the F_1 and F_2 layers cannot be observed. Between the E and F_1 layers there occurs at very irregular times another layer of limited extent, which is sometimes called the E_2 layer (American: "sporadic E layer"); in the summer this is often observed in the morning and in the evening; sometimes this E_2 layer occurs in the evening at all seasons.

The following table gives a survey of the average condition of the ionosphere at noon on a summer day for a mean latitude on the earth:

layer	effective altitude (km)	critical frequency (megacycles/sec)	maximum electron density (number of particles/cm ³)
D	50	<0.4	$<2.5 \times 10^3$
E	120	2.5	10^5
F_1	200	5.0	4×10^5
F_2	300	8.0	10^6

For the sake of comparison it may be noted that from the radio investigation of the ionosphere it appears that the molecular density of the atmosphere at the altitude of maximum ionization of the E layer is of the order of magnitude of 10^{12} molecules per cm^3 , while in the F layer the figure is about 10^{11} .

The question has arisen as to whether there are still other layers present. It is indeed very well possible that we are unable to observe all the layers; a layer can only manifest itself when its ionization is greater than that of all the underlying layers, so that its limiting frequency is higher than the lowest frequency of the waves penetrating through the lower layers. Faint indications of layers other than those mentioned have sometimes been found. It is possible, however, that deceptive phenomena may occur which would appear to indicate the existence of an ionized layer although such a layer does not actually exist. In this connection it is interesting to note that in 1927 Hals received echoes whose echo-time was of the order of magnitude of 10 to 30 sec, which would point to a reflec-

ting layer far beyond the orbit of the moon, where according to Störmer a sphere of charged particles can indeed be expected. Whether this is the correct explanation of the origin of the echoes with long transit time may be open to doubt.

The reason for the complex character of the ionosphere has not yet been determined with certainty. This phenomenon is probably connected with the absorption of different parts of the sun's spectrum at different altitudes in our atmosphere. The upper layers of the ionosphere are probably ionized not only by light but also by charged particles originating in the sun.

Which of the gases of our atmosphere are subject to this ionization is also uncertain. It might be thought that our atmosphere at 200 to 400 km altitude no longer consists mainly of nitrogen, but of lighter gases, for instance hydrogen and helium. A direct indication of the presence of these gases has not, however, been found, although they have often been sought, for example in the spectrum of the northern lights, which occur at the altitudes in question.

One of the bands of the northern lights spectrum could be ascribed to the positive ion of the nitrogen molecule, so that the occurrence of the reaction $N_2 \rightleftharpoons N_2^+ + e$ is thus proved. This process is probably one of the causes of the ionization in the *F* layer.

Daily, annual and other periodical variations

The connection between the position of the sun and the state of the ionosphere has already been referred to in the foregoing. This connection makes it probable that the ionization is for a large part brought about by the ultra-violet radiation of the sun. For the *E* layer this supposition can even be confirmed quantitatively, by calculating the electron density as a function of the position of the sun on the basis of the laws of molecular equilibrium.

The following formula holds for the change in electron density per second:

$$\frac{dN}{dt} = q - aN^2.$$

In this expression q is the number of electrons formed in the ionosphere per second per cm^3 and aN^2 the number which per second and per cm^3 reunite with the positive ions; a is the so-called recombination coefficient. The fact that the velocity of the recombination is determined by the square of N is clear when remembering that the chance of recombination is proportional to the product of the concentrations of the electrons and of the posi-

tive ions. If an equilibrium condition is assumed, i.e. $dN/dt = 0$, the following applies:

$$N = \sqrt{q/a}.$$

The quantity q depends upon the height of the sun, which can be expressed by the angle χ made by the sun's rays with the perpendicular. If one considers, for instance, a horizontal plane in the ionosphere of 1 cm^2 , it is clear that the amount of ultraviolet energy striking that plane is proportional to $\cos \chi$. Thus N is proportional to $(\cos \chi)^{1/2}$, and according to (10) this means that

$$f_{\text{crit}} \text{ is proportional to } (\cos \chi)^{1/4}. \quad (11)$$

This relation is reproduced in *fig. 5* for the sake of comparison with the results of measurements. It is found that the formula for the *E* layer corresponds to the observations carried out during a large

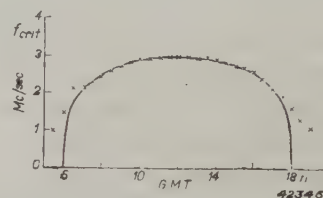


Fig. 5. Daily variation of the critical frequency (in megacycles/sec.) for the *E*-layer (average from March 18 to 23 incl. 1935; taken from Appleton, Proc. Roy. Soc. 162, 451, 1937). The line drawn represents the variation of the quantity $(\cos \chi)^{1/4}$, where χ is the zenith distance of the sun. Except in the hours of night the curve corresponds very well to the variation of the critical frequency. G.M.T. = Greenwich Mean Time.

part of the day. From this it may be concluded that the processes of ionization and recombination in the *E* layer in the daytime are indeed in equilibrium with each other.

Formula (11) does not meet the case so well for the variation for the F_1 layer, and it is certainly not valid for the F_2 layer. In these layers therefore the equilibrium between ionization and recombination is absent, which is understandable considering that in the more rarefied atmosphere the processes take place much more slowly, because of the smaller number of collisions. However, it may also be assumed in explanation that the ionization is determined by phenomena other than the ultraviolet rays. The two causes probably work together.

The other known periodical variations in the state of the ionosphere are likewise connected with the sun. In the course of a year the altitudes and especially the electron densities vary with the seasons (see *fig. 6*). The electron concentration in the *E* layer and especially in the F_1 layer is greatest in summer, which is understandable after what has been said about the influence of the height

of the sun. The F_2 layer behaves differently, the limiting frequency and thus also the electron density reaching a greater value in winter than in summer.

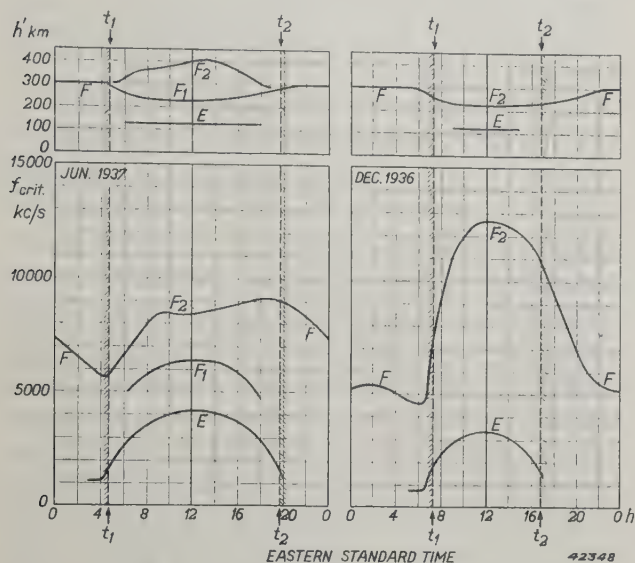


Fig. 6. Daily and seasonal variations of the critical frequency f_{crit} in kc/sec and of the effective altitude h' for the different layers of the ionosphere (borrowed from Smith, Gilliland and Kirby, J. Res. Bur. Stand. 21, 835, 1938). t_1 sunrise; t_2 sunset. Eastern standard time, 75° west longitude (practically the meridian of Philadelphia).

From this it is sometimes concluded that the F_2 layer is sensitive to temperature, becoming diffuse and rising in the summer whereas in winter it becomes denser and falls.

On the basis of his own experiments and those of his collaborators, Elias presumed that the nightly ionization at 350-400 km effective altitude can indeed be distinguished from that in the daytime and that it is to be ascribed to corpuscular radiation, or perhaps partly to electromagnetic radiation of cosmic origin.

The eleven-year period in the number of sunspots observed every year seems to correspond to an equal period in the electron densities of the ionosphere. According to Smith, Gilliland and Kirby, during the years 1932-1937 the electron densities of the E , F_1 and F_2 layers increased with the number of sunspots (see fig. 7).

Further of interest are the changes in the ionosphere during an eclipse of the sun. It has been found that during an eclipse the electron density in the E and F_1 layers decreases sharply, but almost immediately after the eclipse returns to the normal value again. The electron densities of these layers are found to be roughly proportional to the non-eclipsed portion of the sun's disc; that of the F_2 layer behaves differently and varies in a smaller degree. This may be ascribed partly to the recom-

bination coefficient being too low, but on the other hand it seems to confirm the suspicion that the ionization of the F_2 layer is caused partly by a corpuscular radiation.

Finally in 1939 Appleton and Weekes discovered lunar times in the ionization state of the ionosphere. This influence of the moon, however, is very slight.

Disturbances in the ionosphere (fade-outs)

In the year 1935 Dellinger published his detailed investigation of an important discovery by Jouast, Mögel *et al.* At certain moments it is suddenly impossible to receive any shortwave signals passing over the part of earth that is in daylight. This effect occurs so abruptly that it is often thought that something is wrong with the receiving set. The disturbance lasts from about ten minutes to an hour or longer. Dellinger pointed out the simultaneous occurrence of these fade-outs over the whole daylight half of the earth and showed that there was a connection with eruptions in the chromosphere of the sun. The explanation of the fade-outs is to be sought in the fact that when there is an eruption a large amount of ultra-violet radiation is emitted by the erupting spot on the sun, which causes intense ionization in the earth's atmosphere at an altitude of 50 km (D layer), especially through the a line of the Lyman series with a wave-length of 1216 Å. As a result of the relatively high pressure at these altitudes the electrons set in vibration in the field of a radio wave will undergo continual collisions with gas molecules and lose their energy; in consequence the absorption of the radio signals is so great that they disappear and radio communication becomes impossible.

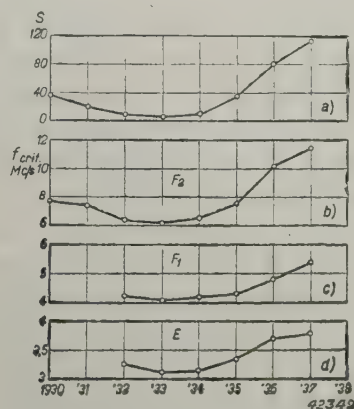


Fig. 7. Comparison of the annual average of the number of sunspots s with the noon values of the critical frequency f_{crit} (in megacycles/sec) for the different layers F_2 , F_1 and E in the corresponding years (from Smith, Gilliland and Kirby, J. Res. Bur. Stand. 21, 835, 1938).

There is some indication that the reception strength of long waves increases during a fade-out, instead of decreasing. Since these waves are reflected by the *D* layer at about 50 km altitude, this would agree with the explanation that fade-out is to be ascribed to an increase of ionization in the *D*-layer.

Disturbances due to *D* layer absorption are also known which do not occur and disappear so suddenly as fade-outs, and which last for a longer time, usually several hours. The decrease in intensity of the signals, too, is usually not so great as in the case of fade-outs.

Disturbances from another cause occur during magnetic storms and upon the appearance of the aurora borealis, which again is found to be connected with the eleven-year period of the sunspots. During the occurrence of magnetic storms and the aurora borealis material particles from the sun penetrate into the ionosphere and there disturb the normal state. Investigation has shown that at a high geographical latitude the electron density of the *E* layer then increases sharply. At lower latitudes the *F* layer seems to be primarily affected; the electron density decreases and the effective altitude rises, thus indicating a diffusion of the layer.

Practical consequence for radio reception

From what has been said in the foregoing a picture can be formed of the influence of the ionosphere in the different frequency regions. In the case of long-wave daytime broadcasting (kilometre waves) the wave is, as it were, enclosed between the *D* layer and the surface of the earth. Owing to the relatively low altitude of the *D* layer the difference between the path travelled by the carrier wave and that travelled by the wave reflected against the *D* layer is usually small compared with the wave length. These waves are therefore not clearly separated from each other but form a single vibration phenomenon which is propagated parallel to the earth's surface. Fading seldom occurs.

A peculiar phenomenon, which may be caused by the ionosphere in the case of long waves, is a sort of cross-modulation, whereby the modulation of a strong transmitter can be transferred to the carrier wave of other transmitters (the so-called Luxemburg effect). This effect was first observed by Butt and Tellegen, each independently, who ascertained that the modulation of the Luxemburg transmitter could also be heard on several other transmitters lying in about the same direction from the receiver, but at greater distances. This

effect is to be ascribed to non-linear phenomena in the ionosphere. Apparently the state of the ionosphere above a powerful transmitter on the long-wave is slightly affected in the rhythm of the modulation, thereby giving rise to a reciprocal action on the radio waves.

Furthermore, on the intermediate wave, between 200 and 600 m fading often occurs, especially at night and particularly in those regions where the intensities of the carrier wave and the wave reflected against the *E* layer are equal. Reception can only be called perfect in the vicinity of the transmitter (radius about 75 km) where the carrier wave is received more strongly than the reflected wave. At a very great distance, where only the reflected wave is received, no fading would be expected, since the cause of the fading phenomena lies in interferences between carrier wave and reflected wave. Nevertheless, even at great distances very marked variations in intensity may occur due to fluctuations in the reflective properties of the ionosphere and also due to interference arising from the fact that when reflected from the ionosphere the signal may reach the receiver along different paths of varying lengths.

During the day the reflected waves of the intermediate region are rather strongly absorbed by the *D* layer.

Wave lengths between 100 and 200 m are usually less suitable for radio communication over long distances. With increasing frequency the carrier wave is more and more strongly damped by absorption in the earth, an effect which is already noticeable at 100-200 m. The reflected wave is weak, since this kind of waves is very much absorbed by the ionosphere. The cause may perhaps be sought partly in the fact that the frequencies of the vibrations of the electrons around the lines of force of the terrestrial magnetic field correspond to the frequencies of this wave region. Resonance is therefore possible, and this is always accompanied by absorption. The wave lengths from 100 to 50 m and from 50 to 10 m are indeed very important for wireless communication. Especially in the latter region, however, the possibility of reception is limited by the occurrence of the dead zone. In the daytime reflection takes place mainly at the *E* layer, while at night the wave penetrates through the *E* layer and is reflected at the *F* layer.

Fading occurs to a very marked degree in this wave-length region. For receivers situated on the edge of the dead zone the average signal strength may vary considerably owing to the expansion or shrinkage of the dead zone. Reception on these

waves depends very much on the activity of the sun. Especially where the trajects pass along the zones of the aurora borealis to earth the reception is very variable.

The strong fading phenomena discussed in the previous section, which are caused by a disturbance in the normal state of the ionosphere, are particularly pronounced in this wave region.

Influence of the earth's magnetic field

In the foregoing it has already been stated that the waves in the region of 100-200 m are strongly absorbed by the ionosphere owing to the terrestrial magnetic field. A closer investigation shows that this is not the only effect of the magnetic field. Under the influence of the earth's field a so-called double refraction occurs, and one speaks, as in the propagation of light in crystals, of the ordinary and extraordinary components. That something of the sort must exist is easily seen when the behaviour at the magnetic equator is studied, where the earth's field is directed horizontally North-South. If at that spot linearly polarized radio waves are sent upwards, whose electrical vector E is also directed North-South, then in the ionosphere the electrons are set in vibration in the direction North-South and they experience in their motion no effect from the earth's field. For such polarized waves the propagation in the ionosphere is independent of the magnetic field and one therefore speaks of the ordinary component. The extraordinary component on the other hand is found when from the magnetic equator radio waves are sent upwards whose electrical vector is directed East-West; the electrons in the ionosphere set in vibration by the radio waves then experience a Lorentz force, with the result that equation (10) for the electron density at which reflection occurs is changed to

$$N = \frac{\pi m}{e^2} f(f \mp f_H) = 1.24 \times 10^{-8} f(f \mp f_H) \quad (12)$$

where $f_H = \frac{e}{m} \frac{H}{2\pi c}$ is the so-called Larmor frequency, i.e.

the frequency at which the electrons can describe circular orbits in the magnetic field, which frequency is equal for all circles. The upper sign holds for $f > f_H$ and the lower for $f < f_H$. For a field strength of about 0.5 Gauss $f_H \approx 1.5$ megacycles/sec., corresponding to a wave length of about 200 m.

When radio waves are sent upwards whose electrical vector E is directed for instance NW-SE, double refraction occurs in the ionosphere. An ordinary component is received which is polarized linearly N-S, and an extraordinary component polarized linearly E-W. When $f > f_H$ it is found from equations (10) and (12) (upper sign) that the extraordinary component is reflected at a smaller electron density than the ordinary component. As a result the ordinary component is received later than the extraordinary one. At the same time the critical frequency for the extraordinary component then lies higher than that for the ordinary one. Thus if the frequency of the signal transmitted is chosen between these two critical frequencies, the extraordinary component is still reflected but not the ordinary component.

Experiments carried out by the Carnegie Institution in Huancayo, Peru (lying on the magnetic equator) fully confirm the above statements. From this it follows in particular that the effective charged particles of the ionosphere are actu-

ally electrons; if the reflection were caused by heavier particles no measurable difference could be expected between the reflection of the ordinary and the extraordinary components.

Experiments carried out at higher magnetic latitudes give a somewhat more complicated result; in that case the reflected waves are polarized elliptically in opposite directions, at the magnetic North pole the ordinary component being polarized circularly to the left and the extraordinary component circularly to the left, while at the magnetic South pole the polarization directions of the two components are just the reverse. Observation of the opposite polarization directions in the Northern and Southern hemispheres clearly confirm these facts.

Significance of the study of the ionosphere for the physics of the sun

In the foregoing it has already been pointed out that the investigation of the ionosphere is closely connected with the physics of the sun, with meteorology and with geophysics. We shall here briefly consider the connection with the physics of the sun.

Some of the ions in the ionosphere have been formed by ionization of nitrogen molecules. This can only be accomplished by radiation of a wave length shorter than 661 Å. If the temperature at the surface of the sun is known, the intensity of the effective ultra-violet radiation can be calculated, considering the sun as a heat radiator. From the distribution of intensity in the visible region a value of at most 6500 °C follows for the temperature of the sun, and from this Saha deduced that in the wave-length region $\lambda < 661$ Å the sun would give off to the earth 10^4 light quanta per cm² per second. Since upon absorption by a nitrogen molecule each quantum releases one electron, in a vertical column with a base of 1 cm² 10^4 electrons are released per second. According to Appleton and Chapman, however, 10^9 to 10^{10} electrons per cm² per second are necessary for the maintenance of the ionization. This result indicates that the intensity of the sun's radiation emitted in the far ultra-violet region is about 10^6 times greater than that of a glowing black body with the temperature of the sun. It is very probable that this extra ultra-violet radiation originates in the hydrogen and helium present on the sun.

The foregoing may serve to show what progress has been made in the radio investigation of the ionosphere. A much deeper insight has been obtained into the physical phenomena taking place in the ionosphere and it is not only radio technics that have profited from this, for the investigation of the ionosphere now also furnishes important material for geophysics, meteorology and astrophysics.

LITERATURE

Among the numerous articles concerning the ionosphere we would mention:

H. R. Mimno, The physics of the Ionosphere, *Rev. Mod. Phys.* **9**, 1, 1937.

W. Grotian, Sonne und Ionosphäre, *Naturw.* **27**, 555 and 569, 1939.

J. H. Dellinger, The role of the Ionosphere in radio wave propagation, *Electr. Eng. Transactions Suppl.* 803, 1939.

K. K. Darrow, Analysis of the Ionosphere, *Bell Syst. Techn. J.* **19**, 455, 1940.

See also:

L. Harang, *Das Polarlicht und die Probleme der höchsten Atmosphärenschichten*, Leipzig 1940.

C. J. Bakker, M. Minnaert, A. Pannekoek, J. Veldkamp, De ver-ultraviolette en de corpusculaire stralingen der zon (The far ultra-violet and the corpuscular radiations of the sun). *Ned. T. Natuurk.* **9**, 209, 1942.

AN EXPERIMENTAL TRANSMITTER FOR ULTRA-SHORT-WAVE RADIO-TELEPHONY WITH FREQUENCY MODULATION

by A. van WEEL.

621.396.5 : 621.396.615.14

A transmitter working with frequency modulation has been developed for an experimental ultra-short-wave radio-telephonic link between the Philips factories at Eindhoven and those at Tilburg, Holland, on a wave length of 90.5 cm for one direction and 99 cm for the other. Modulation with frequencies from 12 to 204 kc/sec, for 48 telephone calls at once, takes place on a carrier wave with a frequency equal to 1/9 of the desired transmitter frequency, with a maximum frequency swing of 67 kc/sec. By frequency multiplication a signal is obtained with the desired transmitter frequency and with a frequency swing 9 times as large. This article points out the advantage of this method and describes how the frequencies have been chosen and the transmitting stages arranged. A brief description is also given of the construction of the transmitter, attention being drawn in particular to the distribution of the circuiting over two separate panels and the simplification of the wiring, both of which have been made possible by the application of a new method of coupling between successive stages, further explained in this article.

An experimental radio-telephony link between the Philips factories in Eindhoven and those at Tilburg¹⁾ has existed for several years. This has been working lately on a wave length of 90.5 cm for one direction and 99 cm for the other. The link is so arranged that the radio installation can function as an entirely automatically acting link in the telephone network. Neither the person phoning nor the operators of the telephone exchange through which the connection passes need to be aware of the fact that the calls are transmitted by wireless instead of by cable.

Since the installation was first set up there have been important developments both in the field of telephony and in short-wave transmitting technique. In telephony there is an increasing tendency to use a single pair of conductors in a cable for the transmission of a large number of calls at the same time by means of a carrier telephone system²⁾. The frequency band of about 3000 c/sec width necessary for transmitting a call is modulated on a specific "carrier", the carriers for the different calls ("channels") lying 4000 c/sec apart (only one side band is used), for instance at 32 kc/sec, 36 kc/sec, 40 kc/sec, etc. It was found desirable to have the radio-telephone link referred to above equipped for such a carrier telephone system, namely for 48 channels, in connection with the available carrier telephone apparatus. The 48 channels, which cover the frequency region from 12 to 204 kc/sec, have to be modulated as a whole on the radio wave, which in this case serves as a "pair

of conductors" in the link. The transmitter and receiver had therefore to be made suitable for this very wide range of modulation frequencies (up to 200 kc/sec).

At that time there was an important development in the field of ultra short waves (< 10 m), viz the gradual superseding of the old method of amplitude modulation by frequency modulation. The advantages of this method led us also to rebuild our ultra-short-wave link for the new method.

All these facts led to a complete reconstruction of the transmitter and receiver of this experimental communication, not only as regards electrical connections but also in the actual construction. Of the old installation only the Yagi directional aerials for transmitter and receiver¹⁾ could be retained unaltered.

In this article the transmitter in its present form will be described, while the receiver will be dealt with in a subsequent article.

The method of modulation

The advantages of frequency modulation over amplitude modulation are twofold: 1) In amplitude modulation of ultra short waves it is almost inevitable that also an undesired frequency modulation occurs; the difficulties created by such a mixed modulation are avoided by employing pure frequency modulation. 2) Frequency modulation, compared with amplitude modulation, gives an appreciable improvement in the ratio between the intensity of the signal and that of the fluctuation noise³⁾.

¹⁾ C. G. von Lindern and G. de Vries: An ultra short wave telephone link between Eindhoven and Tilburg, Philips techn. Rev. 2, 171, 1937.

²⁾ See for example the articles published in this periodical about carrier telephony: Philips techn. Rev. 4, 20, 1939; 6, 325, 1941; 7, 83, 104, 184, 1942.

³⁾ For readers who wish to know more about frequency modulation we refer to the articles by Th. J. Weyers in the preceding numbers of this periodical: Philips techn. Rev. 8, 42 and 89, 1946.

In order to make full use of the second advantage, *i.e.* in order to reduce noise to the lowest possible level, it is necessary that the maximum frequency swing, *i.e.* the largest deviation occurring from the average frequency, should be about ten times as large as the highest modulation frequency to be transmitted. In our case this was not less than 0.2 megacycles/sec, so that the maximum frequency swing had to amount to 2 Mc/sec. Since for the undistorted transmission of a frequency-modulated oscillation, a side band of at least $1\frac{1}{2}$ times the frequency swing on both sides of the carrier frequency has to be transmitted, the transmitter and receiver would have to be adapted for a frequency band with a total width of 6 Mc/sec. Because of the difficulties that would have been involved in getting such a great band width, we confined ourselves to a frequency swing of the emitted signal equal to about three times the highest modulation frequency. In this way one arrives at a band width of 2 Mc/sec, which is well possible in practice, while the ratio between intensity of signal and that of noise is still 14 db greater than it would be with amplitude modulation.

With frequency-modulated transmitters it is customary to apply the actual modulation process to oscillator connections which oscillate at a low frequency. By frequency multiplication, where the average oscillator frequency and the frequency swing are increased in the same proportions, the desired transmitting frequency is then obtained. This method, which we also employed, has the advantage that it is easier to obtain the required proportionality between frequency swing and intensity of the modulated signal. This can be explained as follows.

Modulation is effected by detuning the oscillator circuit with the help of a so-called reactance valve³⁾. This is a normal multigrid valve, between the cathode and anode of which the circuit to be detuned is connected, while part of the circuit voltage is applied to one of the grids with 90° phase displacement. The anode AC is then shifted 90° in phase with respect to the anode AC voltage, in other words the valve acts as a reactance (capacitative or inductive, according as the grid AC voltage is shifted $+90^\circ$ or -90° in phase). To take a specific case, let us assume that the tube acts as a capacity, C_1 . The reactance $1/\omega C_1$ is equal to the ratio of the amplitudes of anode AC voltage and anode AC, while the frequency ω at which the circuit connected oscillates is determined by C_1 together with the remaining circuit capacity C_0 . C_1 is now varied by applying the modulating signal voltage to a second

grid of the valve, whereby the slope, and with it the amplitude of the anode AC, is changed in the rhythm of that voltage. The detuning, $\Delta\omega$, of the circuit obtained with respect to the frequency ω_0 , with $C_1 = 0$ is given by

$$\Delta\omega = \frac{\omega_0 C_1}{2C_0}.$$

We will now consider the influence of the oscillator frequency to be chosen, ω_0 . In connection with the linearity of the modulation, only a certain change in slope of the reactance valve is permissible, namely such that there is no deviation from that section of the valve characteristic where the slope is by sufficient approximation proportional to the grid voltage. Corresponding to the permissible maximum slope, which depends only on the properties of the valve and not on the oscillator frequency ω_0 , is a certain maximum anode AC, thus — because the anode AC voltage is fixed — a certain maximum value of the reactance $1/\omega_0 C_1$, likewise independent of ω_0 . The maximum detuning to be obtained $\Delta\omega$, which according to the formula is proportional to the greatest value of $\omega_0 C_1$, is therefore independent of the frequency ω_0 , because C_0 may also be considered as a constant. In order to make the detuning by the reactance valve as great as possible, C_0 will be kept as small as possible, and thus will be limited to the unavoidable capacities of valve and wiring. The choice of ω_0 is then realized with the self-induction of the circuit.

Since, therefore, the same absolute frequency sweep can be obtained with a low oscillator frequency as with a high one, if in the manner described one begins with a low oscillator frequency and a correspondingly small frequency sweep, it will actually be easier to ensure the necessary linearity of the modulation. We shall revert later to the exact choice of oscillator frequency.

The connections

The connections are in push-pull arrangement for various reasons: variations in the feeding voltage are much less manifest as undesired modulation, since they act on the two halves of the modulator in the same phase instead of in counter-phase; furthermore one avoids the strong high-frequency currents which otherwise flow in the earth connections and therefore (since all the points to be earthed cannot be connected to the same point of the chassis) also through the chassis, causing all kinds of undesired couplings, etc; finally, by using push-pull amplifier valves several difficulties occurring at very high voltages are diminished, because,

among other reasons, the influence of the self-induction of the cathode feeding connections is much smaller⁴). In *fig. 1* a block diagram is given of the transmitter connections. We shall first consider only that part drawn with heavy lines. The low-frequency signal from the telephone exchange (the term "low-frequency" is here used in a relative sense, the frequencies of this signal being as high as about 200 kc/sec.) is led to two reactance valves which influence the frequency of the push-pull oscillator. The signal obtained from the oscillator, i.e. the frequency-modulated oscillator voltage, is

explained above that the oscillator frequency must be chosen as low as possible. From the frequency values indicated in *fig. 1* in the different stages it may be seen that the choice fell upon 36.9 Mc/sec. This still seems relatively high, but it must be taken into account that the receiver for the same connection is situated close to the transmitter. The receiver works on the superheterodyne principle with an intermediate frequency of 18 Mc/sec. In order to prevent interferences in the reception it is necessary that neither the oscillator frequency of the transmitter nor any harmonics of it shall fall

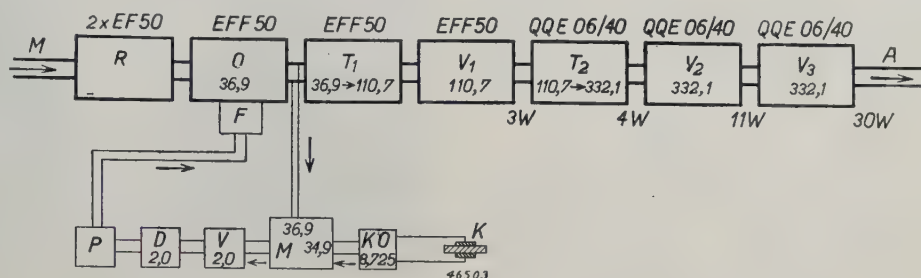


Fig. 1. Block diagram of the transmitter. The modulating low-frequency voltage is applied at *M*, *R* reactance valves, *O* oscillator, *T*₁, *T*₂ frequency-triplicator stages, *V*₁-*V*₃ amplifier stages. At *A* the frequency-modulated output voltage is fed to the aerial. The numbers in the blocks indicate the frequencies in Mc/sec at which the stages work; above each block the type of valve used for that stage is given. The part of the diagram drawn with thin lines serves to keep the average oscillator frequency constant. This is explained later (see *fig. 3*).

led to a push-pull valve, connected as a frequency triplicator. Then the signal is amplified and conducted to another frequency triplicator. The signal here reaches its final frequency, after which it is further amplified in two end stages and sent to the aerial.

The choice of oscillator frequency

The reason for the multiplication of the frequency in stages by a factor of three lies in the nature of the connections. In push-pull connections the output voltage can contain, in principle, only the odd harmonics of the frequency of the grid AC voltage applied, so that by filtering out the harmonics in question a frequency multiplication by a factor 3, 5 or 7, etc. can be obtained. For the sake of efficient output the lowest factor has been chosen. Since the transmitting frequency, if we confine ourselves for the moment to one single call direction, was fixed at 332.1 Mc/sec (90.5 cm wave length) the possible choices of oscillator frequency were 332.1 - 110.7 - 36.9 - 12.3 - 4.1 Mc/sec, etc. We have

in the intermediate-frequency band of the receiver, which band, according to the above figures, must extend from 17 to 19 Mc/sec. An oscillator frequency of 12.3 or 4.1 Mc/sec. would, it is true, also answer this condition, but the transmitting apparatus in question had to be so designed that later on, if necessary, the connections built up with the same stages could also be used for telephone links on other wavelengths between about 1.50 and 0.90 m, simply by changing slightly the oscillator frequency. In order to avoid once for all the danger of interferences in the receiver it was therefore decided to place the oscillator frequency above the intermediate-frequency band of the receiver. At the same time of course the possibility had to be considered of choosing that intermediate frequency itself lower; in the discussion of the receiver it will be shown, why this was not done.

It is perhaps advisable to stress the point that the requirement mentioned above, that for a favourable ratio between signal and noise the frequency sweep must be several times as large as the highest modulation frequency occurring, is only applicable for the signal emitted by the aerial. In our case, as a consequence of the frequency multiplication, the maximum frequency swing in the oscillator stage amounts to only 1/9

⁴) M. J. O. Strutt and A. van der Ziel: A new push-pull amplifier valve for decimetre waves, Philips techn. Rev. 5, 172, 1940.

of the final value, i.e. $1/9 \times 3 \times 200 = 67$ kc/sec. If a modulating signal with the frequency of 200 kc/sec and with the largest permissible amplitude is applied to the oscillator, the oscillator delivers an oscillation with a frequency fluctuating 200 000 times per second between 36.833 and 36.967 Mc/sec. In the case of the frequency multiplication here employed the rhythm of the fluctuation, i.e. the modulation frequency, naturally remains unaltered: the following stage delivers an oscillation with a frequency fluctuating 200 000 times per second between 110.5 and 110.9 Mc/sec, etc.

The valves used

The valves used are indicated over the blocks in the diagram of fig. 1. Except for those in the three last stages they are all normal receiving valves with only a low energy dissipation so that a compact assembly is possible. The last stages must of course have transmitting valves in order to produce the power required for transmission.

Nevertheless, the energy amplification in the last two stages is only slight, as may be seen from the wattage figures given in fig. 1. This is due to the high frequency at which these amplifier stages have to function. It would seem obvious to ask why the once amplified signal of 110.7 Mc/sec is not first amplified further to the desired final level and then given the necessary frequency amplification in the last stage. This, however, is impossible, because if that were done very high AC voltages (of the order of 500 Volts) would have to be applied to the grid of the last valve; in order to function as frequency multiplier the valve must work in class C, thus with very high negative grid bias. It has been found in practice that with the very small distance between grid and cathode in these short-wave tubes the high voltages mentioned lead to disturbances: breakdown may occur or the insulation between grid and cathode may be damaged (especially at very high frequencies).

Fig. 2 is a reproduction of two X-ray photographs of the type of short-wave transmitting valve used, the QQE 06/40. It is a double tetrode in which the two balanced systems have a common indirectly heated cathode and a common screen grid. The power on a wave length of 3 m is 40 W; at 1 m about 30 W.

Maintaining the constancy of the oscillator frequency

The part of the diagram in fig. 1 which is drawn with thin lines serves to keep the average oscillator frequency constant; it is given again separately in fig 3.

The usual method of synchronizing the oscillator vibration directly with the characteristic oscillation of a quartz crystal or a harmonic of the same

could not be employed in our case since the oscillator vibration is already frequency-modulated upon its formation, and consequently when keeping the

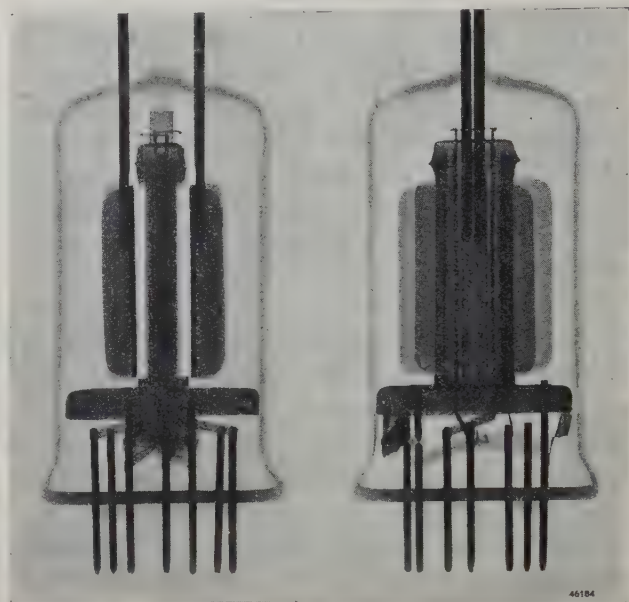


Fig. 2. X-ray photographs taken in two mutually perpendicular directions of the ultra-short-wave push-pull transmitter valve, type QQE 06/40, used in the three last stages of the transmitter. The cathode connection is made very short thanks to the two electrode systems having a common indirectly-heated cathode.

oscillator frequency constant a certain margin has to be left for the frequency swing. For that reason the following method was chosen.

A small part of the output voltage of the oscillator stage, of which the average frequency f_0 must amount nominally to 36.9 Mc/sec, is tapped off and

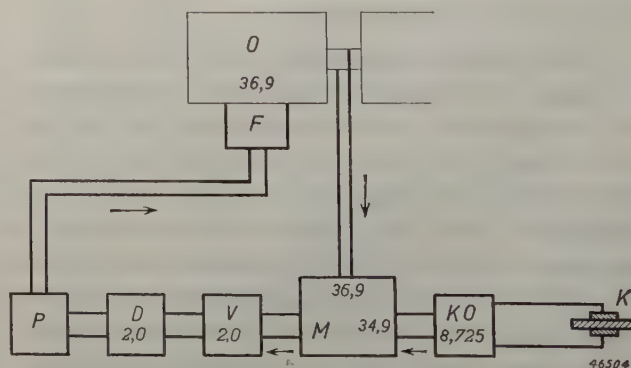


Fig. 3. Block diagram of the connections for keeping the average oscillator frequency constant. K quartz crystal which keeps the frequency 8.725 Mc/sec of the crystal oscillator KO constant, M mixing stage in which the fourth harmonic of this frequency (34.9 Mc/sec) is mixed with the frequency $f_0 \approx 36.9$ Mc/sec of the transmitter oscillator O, V amplifier, D discriminator, P pentode, F coil with ferromagnetic core coupled magnetically with the circuit self-induction of the oscillator O. The numbers in the blocks indicate the frequencies in Mc/sec.

in a mixing valve mixed with an AC voltage of the very constant frequency of 34.9 Mc/sec from a vibrating crystal. The output voltage of this mixing stage has an average frequency $f_1 = f_0 - 34.9$ Mc/sec. If f_0 is exactly 36.9 Mc/sec $f_1 = 2$ Mc/sec; if f_0 differs slightly from 36.9 Mc/sec f_1 exhibits the same absolute difference from 2 Mc/sec. Furthermore, the "intermediate frequency" f_1 is of course frequency-modulated with the telephone frequencies in the same way as the high frequency f_0 applied. This output voltage of the mixing stage is now amplified to a discriminator connection. This produces a DC voltage proportional to the difference between the average frequency f_1 of the applied voltage and the fixed frequency $f_d = 2$ Mc/sec at which the discriminator is set. An AC voltage is superposed on

temperature, etc. The fact that this forms no obstacle to the regulating action of the whole is due to the discriminator reacting to the absolute changes of the oscillator frequency. Even if the discriminator frequency f_d should drift proportionally just as much as the oscillator frequency f_0 does with no regulation, the absolute variations of the latter are still $36.9/2 \approx 18$ times as large as the drifts of f_d .

The coupling between the successive valves

For the coupling together of the various tripliator and amplifier stages a new method has been employed which offers important advantages. In order to explain the particulars of this coupling, let us consider *figs. 4a* and *b* in which a coupling

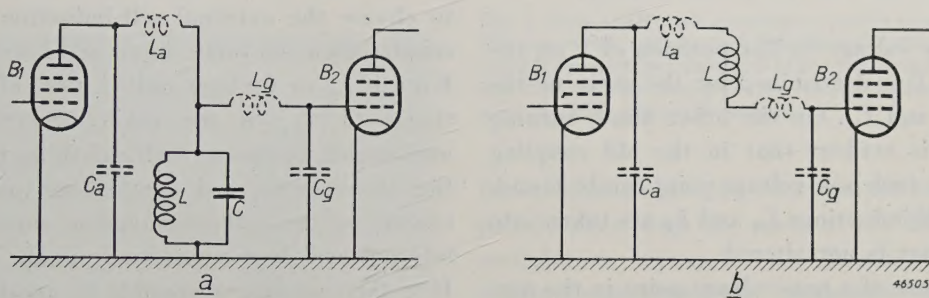


Fig. 4. a) Diagram of the usual method of coupling two amplifier valves B_1 and B_2 ; L - C tuned parallel circuit. b) New method of coupling. C_a , C_g , and L_a , L_g are the internal capacities and self-inductions of the valves involved in the coupling. The elements necessary for determining the DC voltage situation are not drawn.

the DC voltage, namely the original modulation voltage, which, however, is suppressed by a low-pass filter. The DC voltage is now applied to the grid of a pentode. The anode current of this valve, which is thus proportional to the absolute deviation of the average oscillator frequency from the prescribed value of 36.9 Mc/sec, flows through a coil wound around a core of ferromagnetic material⁵⁾ magnetically coupled with the selfinduction coil of the oscillator circuit. Owing to the fact that the anode current changes the premagnetization and thus the permeability of the coil core, it affects the self-induction of the oscillator coil in the sense that the change in the average oscillator frequency, to which the anode current is proportional, is opposed. In this way it proved possible to reduce the drift of this frequency by a factor 30.

The "fixed" frequency f_d at which the discriminator is set is, of course, not actually fixed either, but subject to some variation due to the drift of the elements of the discriminator connections with

according to the usual method and one according to the new method are shown side by side. For the sake of simplicity ordinary (not push-pull) stages are assumed and all elements only of importance for the DC voltage situation are omitted. In the ordinary coupling the anode of the first and the grid of the second valves are connected directly with each other, while between this point and earth a tuned parallel circuit is connected (the internal self-inductions L_a and L_g of the valves, which are indicated with dotted lines in *figs. 4a* and *b*, may for the present be disregarded). In the new coupling, on the other hand, the tuned circuit is formed by the self-induction L and the connection in series of the internal valve capacities C_a and C_g . In *figs. 5a* and *b* the situation in the two cases is shown still more concisely. Let us first consider *fig. 5b*. The high-frequency voltage between the points A and B is mainly determined by the oscillation of the circuit, but the potentials of these two points fluctuate in opposite phase with respect to earth due to the obtrusion of earth potential at point O in the middle of the circuit capacity. From this it

⁵⁾ It is necessary to use a ferromagnetic material the losses of which are sufficiently small, even at high frequencies.

follows that also in the middle of the circuit self-induction there must be a point (P) which remains at earth potential during oscillation and thus has no

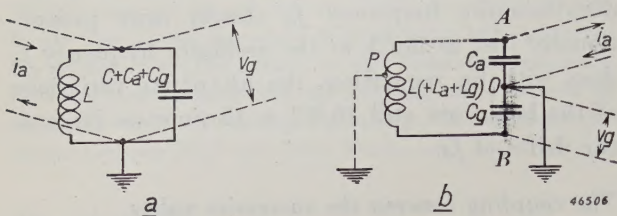


Fig. 5. a) More detailed diagram of the usual coupling circuit. The anode current i_a of the first valve and the grid voltage v_g for the second valve are respectively applied to and taken from the same pair of terminals.

b) The same for the new method of coupling. Here i_a is applied to the "terminals" of C_a , while v_g is taken from the terminals of C_g . Since the point O lying in the middle of the circuit capacity remains at earth potential, there is also a point P on the self-induction with the same property.

high-frequency voltage ⁶⁾. The position of P on the self-induction L is determined by the ratio of the capacities C_a and C_g . On the other hand, turning to fig. 5a, it is evident that in the old coupling connections no such non-voltage point can be found. Even if the self-inductions L_a and L_g are taken into account this fact is not altered.

The occurrence of a non-voltage point in the new method of coupling offers the possibility of dividing the transmitter constructionally into two parts at a non-voltage point between two valves and assembling the two parts, for example, in separate panels. The connection between the two panels then carries no high-frequency voltage (high-frequency current does, however, flow through it) and therefore no undesired couplings or radiation can occur. If such a division were made at a point which was not voltage-free, the connection would have to be shielded against the effects mentioned, and a large extra circuit capacity would thereby be introduced, in general resulting in a loss of amplification. For this reason it is very difficult to divide the connections when the old method of coupling is used.

Another advantage of the new method of coupling is of particular importance at very high frequencies. Since in this case the capacity and/or self-induction of the coupling circuit must become very small, the contributing, unavoidable capacities and self-inductions of the valves begin to play an important part. Fig. 4a cannot then be reduced to the simple situation of fig. 5a, because of the presence of L_a and L_g , which can no longer be ignored. The ordi-

nary connections are now in principle more complicated than a simple $L-C$ circuit, and in practice it proves difficult to obtain sufficient amplification with them. This is understandable when it is borne in mind that at very high frequencies L_g may already be approximately in resonance with C_g . The tuned parallel circuit $L-C$, whose task it is to furnish a high impedance for the circuit frequency, is then as it were short-circuited by the low impedance of the series circuit L_g-C_g . Measures for overcoming these difficulties are known but they only result in making the circuit still more complicated, more extensive and more difficult to manipulate. When we compare this method with the new coupling according to fig. 4b, we see that the presence of L_a and L_g does not alter anything in the principle of the connections and only makes it necessary to choose the external self-induction L somewhat smaller than the total circuit self-induction desired. Even if L_g or perhaps only a part of it is of itself tuned to C_g , in the main everything remains unchanged. It means really nothing else than that the above-mentioned non-voltage point then lies exactly at the valve terminal or somewhere in the internal grid feed connection, as the case may be. It is then no longer possible to divide the connections at that point, but the action of the amplifier stage is not at all affected. It would only become troublesome if a negative value should be required for the external self-induction L . But since C_a is usually much smaller than C_g , the self-induction necessary to tune C_a , i.e. the self-induction between the anode and the non-voltage point, is generally large enough to cover, in addition to the self-induction L_a of the internal anode feed connection, a positive external self-induction as well.

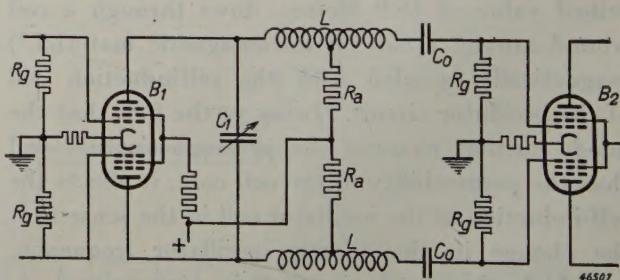


Fig. 6. Example of a push-pull amplifier stage according to the new method of coupling. Between the anodes of the push-pull valve B_1 and the control grids of the valve B_2 two self-inductions L are connected. The capacities C_0 only have the function of separating the DC voltage positions of anodes and grids. The circuit is tuned with the variable capacity C_1 . The anode DC is supplied through the resistances R_a ; R_a may not therefore be chosen very large; in spite of this in order not to obtain any undesired damping of the circuit these resistances are connected at the non-voltage points. There is no objection to choosing large grid leakage resistances R_g .

⁶⁾ Strictly speaking this is not true. If the losses of the circuit elements are taken into account it is found that there is a certain residual voltage. This is very small, however, compared with the circuit voltage.

Fig. 6 shows the complete connections of a push-pull amplifier stage with the new coupling method. Several details are explained in the text of the figure.

Construction

If a transmitter is to serve as part of a carrier telephone system it is desirable that its external form should be adapted to that of the carrier telephone apparatus. In the case in question the transmitter had to be housed in a rack of certain dimensions with the components mounted on panels of a certain length and width which could be slid into the rack from front and rear. Similarity in external appearance, however, can never be complete. In the case of the carrier telephone apparatus

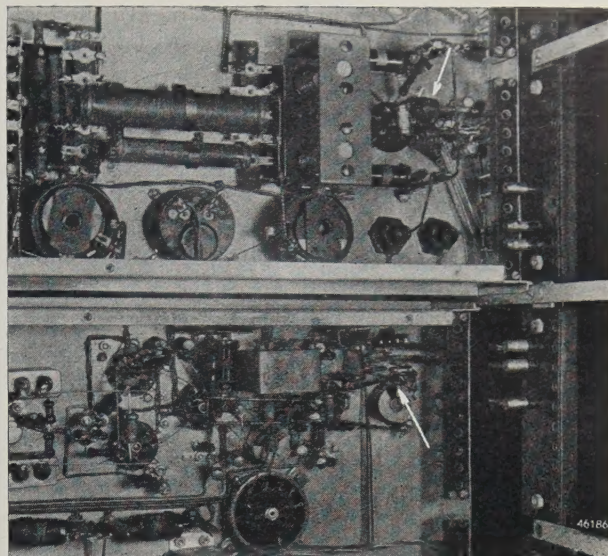


Fig. 8. Part of the two transmitter panels seen from the rear. The two small coils in each panel on the extreme right (indicated by arrows) are the parts of the two coupling coils divided into two. In the lowest drawer, pulled halfway out, may be seen the plug pins and sockets.

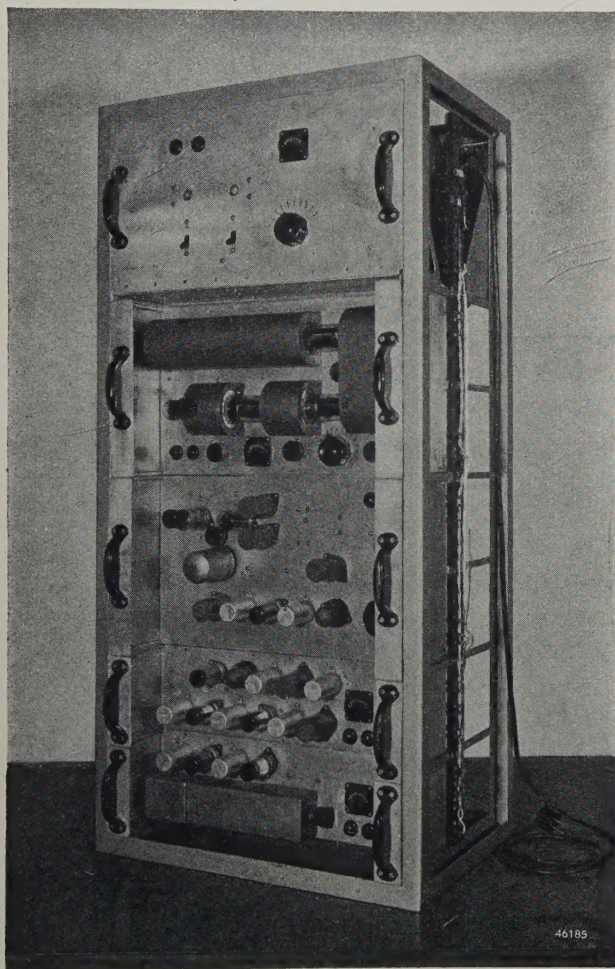


Fig. 7. The transmitter and receiver assembled in a rack. From top to bottom: a panel (of double depth) containing the high tension supply unit, two panels with the transmitter, two panels with the receiver. All the panels are constructed as sliding drawers. On the front plate of the upper transmitter panel may be seen the valves of the three last stages, within, between the coupling connections, shielded by caps. On each panel is a meter with which the cathode currents of all the valves can be checked by means of a switch. In the open spaces at the back of the rack identical drawers with a transmitter and receiver can be inserted to serve as reserves.

the available space in every panel is almost completely filled with the parts of the different filters, modulators, repeaters, etc. Because of the much higher frequencies, the components of the transmitter, however, must in general be spread out over a certain area in order to provide sufficient mutual intervals to prevent undesired couplings. One may speak here of surface assembly in contrast to the volume assembly in the apparatus for carrier telephony. In this way one arrives at the construction shown in fig. 7. The components are mounted side by side in each panel in a single plane, a "front plate". This lies fairly deep in the rack, so that the different valves, coil cans, Lecher systems, etc. situated on the front of the plate are well protected against undesired contact or shocks without it being necessary to place a cover plate in front of the rack, which would hinder the dissipation of the heat given off by the valves to the air.

The uppermost panel of the installation is actually double, occupying the entire depth of the rack in order to offer space for the high tension supply unit mounted therein. This is placed at the top in order that the rather large amount of heat developed in it shall not cause any difficulties in the transmitter part proper. The transmitter also contains too many components to be housed on one normal panel. As a consequence of the above-mentioned method of coupling between successive stages the transmitter could without difficulty be divided and housed in two separate panels one above the

other. This can be seen in fig. 7 underneath the high tension supply unit. Fig. 8 shows the rear of the transmitter panels with the parts of the two coupling coils (push-pull connections, see fig. 6) in each panel. Below the transmitter panels is the corresponding receiver, also divided between two panels. On the other side of the rack two exactly identical panels with a complete transmitter and receiver can be inserted to serve as reserves, and in case of a breakdown these can be switched on by a single touch of the hand or entirely automatically.

Upon inserting a panel all connections of the various signal and feeding voltages are automati-

cally made by means of plug pins in the rack and sockets in the panels. The lead wires to the plug pins are led out at the side, where the wiring harness for the mutual connection of the panels is mounted. The details can be seen in figs. 7 and 8. On each panel a measuring instrument is mounted on which the cathode current of each valve can be checked by means of a switch. By this means it is possible to localize the cause of any interruptions quickly, while the ageing of the valves can also easily be ascertained in good time. This is necessary because the same requirements as to reliability are made of the transmitting apparatus as are made of the telephone apparatus itself.